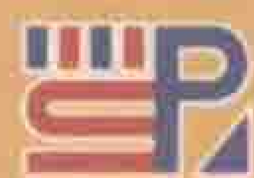
生命科学名著

蛋白质物理（原书第四版）

Protein Physics (Fouth Edition)

[俄] A. V. 芬克尔施泰因(Alexei V. Finkelstein)
O. B. 普季岑(Oleg B. Ptitsyn) 著
李安邦 译



科学出版社



生命科学名著·典藏版

呈现国际学术精华，引进国际经典名著

- 细胞生物学精要（原书第三版）
- 分子生物学（原书第五版）
- 基因的分子生物学（原书第七版）
- Brock微生物生物学（上下册）
- 遗传学：基因和基因组分析（第八版）
- 细菌分子遗传学（原书第五版）
- 神经生物学：从神经元到脑（原书第5版）
- 人类分子遗传学
- 衰老分子生物学
- 发育的原理（影印版）
- 糖生物学概述（原书第三版）
- 结构生物学：从原子到生命
- Brock微生物生物学（影印）（原书第十版）
- 生物化学——基础理论与临床（原书第六版）
- 细胞生理学手册——膜生物物理学精要（下册）（原书第四版）
- 细胞生理学手册——膜生物物理学精要（上册）（原书第四版）
- 植物生物学
- 表观遗传学
- 蛋白质物理
- 进化
- 细胞
- 藻类学
- 基因组3
- 森林遗传学
- 基因组学概论
- 癌生物学
- Lewin 基因X
- 糖生物学基础
- 微生物基因组
- 兽医微生物学



科学出版中心　生物分社
联系电话：010-64012501
E-mail：lifescience@mail.sciencep.com



赛拉艾芙
生命科学订阅号



本书彩图及更多
信息请扫码

www.sciencep.com



销售分类建议：生物化学／生物物理

定价（全套）：4500.00 元

生命科学名著·典藏版

Protein Physics
蛋白质物理

（原书第四版）

〔俄〕A.V.芬克尔施泰因（Alexei V. Finkelstein）
O.B.普季岑（Oleg B. Ptitsyn） 著

李安邦 译

本书的出版得到了国家自然科学基金(项目号:11175068)的支持

科学出版社
北京

图字：01-2013-2616 号

内 容 简 介

本系列丛书均选择生命科学领域经久不衰的经典名著，作者均为国际一流专家，堪称各个专业领域的国际第一书。每一本书的更新都紧跟学科发展，更加适合当前的学习和研究。

本丛书包括《癌生物学》、《分子生物学》、《神经生物学：从神经元到脑》、《表观遗传学》、《基因的分子生物学（第七版）》、《细胞生物学精要（原书第三版）》、《结构生物学：从原子到生命》等 30 本经典著作。

图书在版编目（CIP）数据

生命科学名著：典藏版/（美）温伯格（Weinberg, R. A.）等编著；詹启敏等译. —北京：科学出版社, 2016
ISBN 978-7-03-047485-8

Ⅰ.①生… Ⅱ.①温…②詹… Ⅲ. ①生命科学–研究 Ⅳ.①Q1-0

中国版本图书馆 CIP 数据核字(2016)第 043879 号

责任编辑：王 静 李 悦
责任印制：张 伟 / 封面设计：刘新新

科学出版社 出版
北京东黄城根北街 16 号
邮政编码：100717
http://www.sciencep.com
北京厚诚则铭印刷科技有限公司 印刷
科学出版社发行 各地新华书店经销
*
2016 年 7 月第 一 版 开本：787×1092 1/16
2016 年 7 月第一次印刷 印张：1431 3/4
字数：33 950 000
定价：4500.00 元

(如有印装质量问题，我社负责调换)

中文版序

我非常高兴，在我的朋友李安邦——他在莫斯科大学听过我的蛋白质物理课——的巨大努力之下，蛋白质物理课可以进入中国的科学讲堂，世界上最大的讲堂。在这些讲义翻译为中文之前的全新编辑过程中，我大大地增加了供读者练习的习题的数量，当然，还尽力把在前一俄语版出版以来最近7年蛋白质物理学中所出现的所有最突出的成果包括在本书中。

A. V. 芬克尔施泰因(Alexei V. Finkelstein)
2012年12月

前言(第三版俄文版)

新版《蛋白质物理》在 Oleg B. Ptitsyn(奥列格·普季岑)诞辰 75 周年和他过早地逝世 5 周年之际付印了。

在筹划这个版本时,我首先改正了暴露出的文字错误、某些表格和图片中的不准确或不明确之处。然后,我在第 11 讲和第 18 ~ 22 讲中引入了新内容。之后,在 А. К. Цатурян 和 С. Ю. Бершицкий 的帮助下(我对他们深表感谢),我重新编写了第 25 讲中关于肌肉工作的部分。最后,我在新版中增加了一定数量供读者自主练习的习题。

我感谢《蛋白质物理》俄文版和英文版的读者们以及听我的讲座的听众们,他们向我说出了或寄给我他们的意见,使得我对这些讲义进行修正。我特别要感谢以下向我提出意见和提供附加材料的以下人士:Д. Н. Иванков, Е. С. Надеждина, В. Е. Бычкова, О. В. Галзитска, С. О. Гарбузинский, Н. С. Богатырева, Г. В. Семисотнов, Н. Ю. Марченко, Е. Н. Барышникова, М. Г. Шарапов, Л. Б. Переяславц, Ф.К. Гиоева, А. Я. Бадретдиний, А. А. Климов, Д. А. Климов, В. Е. Финкульштейн, А. Александреск, А. Кистер, Д. Бейкур 和 К. Плакско。

我感谢 А. А. Веденов, А. М. Гутин, Е. И. Шахнович, Б. А. Реве, М. Д. Франк-Каменецкий, А. М. Дыхне, С. Чотиа, М. С. Гельфанд, И. М. Гельфанд, К. Добсон, А. Фершт, Г. Фринд, А. Трамонтано и А. Леск 对讲义中涉及的问题所进行的讨论。

我感谢 М. Г. Никитина 在本书新版本出版工作过程中给予的帮助,感谢我的妻子 М. С. Вильнер-Мармер,没有她长期的支持、关注和协助,本书的准备工作不可能最终完成。

最后,我感谢莫斯科市津贴、ISSEP 基金、俄罗斯科学院蛋白质研究所学术中心和 Algodayn 公司对我的教学活动的支持,感谢俄罗斯科学院、俄罗斯科学院分子和细胞生物学计划、"领先科学流派"计划,俄罗斯基础研究基金[①]和霍华德·休斯医学研究所对我们的(由 Oleg B. Ptitsyn 创建的)实验室的长期支持。

A. V. 芬克尔施泰因(Alexei V. Finkelstein)
2004

①相当于中国的国家自然科学基金

序(第二版英文版)

1967 年 6 月,Oleg Ptitsyn(奥列格·普季岑)成为普希洛新成立的蛋白质研究所的蛋白质物理实验室的负责人。三个月之后,Alexei V. Finkelstein(阿列克谢·V. 芬克尔施泰因)参与到他的工作中来——最初是作为研究生,后来是作为同事。由于受到高分子聚合物物理学的俄罗斯学派的强烈影响,他们研究蛋白质的方法与西方的常用方法并不相同。该流派中最著名的成员之一,Michael Volkenstein,曾经是 Ptitsyn 的博士生导师。Oleg Ptitsyn 和 Alexei V. Finkelstein 一起在普希洛创建了一座世界上杰出的蛋白质物理和化学研究中心。

他们在某些领域特别是关于蛋白质折叠的工作,或是直接通过他们的科学论文,或是间接地通过他们的两位学生——在美国的 Eugene I. Shakhnovich①,在英国的 Alexey G. Murzin②——对其工作的辛勤耕耘,已经在欧洲、印度和美国广为人知。但是,很显然,当 Finkelstein 或 Ptitsyn 与同事们交谈时,他们的工作范围远远超过了西方学者所熟知的。我们会发现,我们当前所关注的一些基础问题已经早已被他们思考过,并且已经给出了一些能得出答案的优美的计算。以前,他们没有抽出时间来出版他们的著作。现在,为了让他们的全部成就不只是让他们的朋友和莫斯科大学的学生们可以接触到,Finkelstein 在他和 Ptitsyn 对他们的学生所做的讲座的基础上写成了此书。

无论是从论述范围的宽广性还是从分析和思维一致的严谨性来看,此书都是绝世之作(*tour de force*)。考虑到蛋白质物理和化学所涉及的内容,我毫不怀疑,任何人,无论是他们的学生或高级工作者,都会从中找到新的、有用的且重要的思想、解释和信息。

Cyrus Chothia③

英国剑桥大学 MRC 分子生物学实验室④

---

①Eugene I. Shakhnovich,美国哈佛大学化学与化学生物学系教授。他通过使用统计物理方法等多种方法,对蛋白质折叠问题作出了基础性的重要贡献。例如,他发现了多肽序列成为类蛋白质的判据,在动力学上他发现了成核机制和折叠核。参见本书第 16、18 和 21 讲。

②Alexey G. Murzin,英国剑桥 MRC 分子生物学实验室教授,蛋白质结构分类法 SCOP 的发明者。SCOP 提供了对已知蛋白质的结构之间的相互关系的详尽说明,已成为结构基因组学中对新蛋白质进行功能性注释的基础资源。由于 Alexey Murzin 在 SCOP 分类法和数据库中的核心作用,他可能是世界上唯一对所有蛋白质结构都了然于胸的人。参见本书第 14、15 和 23 讲。

③Cyrus Chothia,英国皇家科学院院士,英国剑桥 MRC 分子生物学实验室教授。他和 Michael Levitt 共同提出了对蛋白质的"α 型、β 型、α+β 型和 α/β 型"结构分类,与 Alexey Murzin 合作建设 SCOP 数据库,与 Julian Gough 共同建立了 SUPERFAMILY 数据库。

④英国剑桥大学 MRC 分子生物学实验室是世界著名的分子生物学研究机构,是现代分子生物学的发祥地之一。

前言(第一版俄文版)

呈现于你们面前的这个讲义是关于蛋白质物理学的,即关于蛋白质分子的最基本的结构、自组织和功能问题。

本讲义是基于我们(开始时是 Oleg B. Ptitsyn,后来是 Alexei V. Finkelstein)最初在莫斯科物理技术学院、后来在普希洛①大学和莫斯科大学②普希洛分校所做的讲座。讲座最初是针对物理学家的,然后主要是针对生物学家的(部分地对于化学家)。因此在最近这个讲义不仅进行了很大的更新(科学不是停滞不动的),而且进行了彻底修订,以适应生物学课堂。

因此,在大家面前看到的,是课程讲义手稿,而不是专著或教科书。这里难免有重复之处(主要是图片的重复)。因为在实际讲课中不可能说"请看前一讲中的图 2 和公式 3";不过,我已经尽量把这些重复之处减至最少。

我们深深地感谢在本书中用到的我们的工作的所有共同作者,以及阅读本书手稿、讨论并给出了大量有益评论的 А. И. Четверин, Д. И. Харакоз, В. У. Бычкова, Ю. В. Митин, А. С. Спирин, В. А. Колб, М. А. Ройтберг, В. В. Великов, А. В. Ефимов, И. Г. Птицина, Г. И. Гительзов, Д. С. Рыкунов, Г. В. Семисотнов, А. В. Скугарев и Д. Н. Иванков。

我们感谢在我们俄罗斯科学院蛋白质研究所的所有同事,特别是在我们的物理实验室里工作过和正在工作的。没有与他们的持续不断的交流(科学方面和简单的私人方面的),本书就不可能写成。

我们感谢本讲义的所有听众:我们的学生,以及国际学习班和俄罗斯学习班的参与者,他们的问题和意见让作者们更明确地,首先对于我们自己,处理本讲义中涉及的许多问题。

我们要特别感谢 Г. А. Морозов, Д. С. Рыкунов, А. В. Скугарев, Д. Н. Иванков, М. Ю. Лобанов, Н. Ю. Марченко, И. В. Соколовской, А. А. Шумилин,

---

①普希洛科学中心(Pushchino Scientific Centre, http: //www. psn. ru)是位于莫斯科南部 110km 处的一个科学城,建立于 1963 年,现有十多个研究所,主要从事物理化学生物技术领域的研究。目前有 3200 多人在那里工作,其中 1300 多人为科研人员,其中 850 人有副博士或科学博士学位(在苏联/俄罗斯,"副博士"相当于国际上通行的"博士",而"科学博士"是科学工作者通过多年的科研工作之后进行论文答辩所能获得的最高学位)。

②罗蒙诺索夫国立莫斯科大学是苏联/俄罗斯最好的综合性大学,列居世界排名前十名内。1755 年由教育家 M. B. 罗蒙诺索夫倡议并创办,在 20 世纪 50 年代扩建。该校目前有超过 8600 位教师,其中一半有副博士或科学博士学位,125 位是俄罗斯科学院院士,在校学生约 3 万人。

М. Г. Дашкувич, Т. Ю. Сальникова, Н. С. Богатырева, А. Э. Жумаев 和 А. А. Финкульшнейн 等在讲义的图片设计和文字排版方面给予的帮助，也要特别感谢俄罗斯基础研究基金对本书出版给予的资助。

事先的半哲学性的注解

关于“讲课人”。所有讲义是按照它们在教室里讲课时那样写作的：即以“讲课人”这个人物的第一人称方式进行。

关于“课堂里的声音（*inner voice*）**”**。“讲课人”身份是两个共同作者（Oleg B. Ptitsyn 和 Alexei V. Finkelstein）合二为一而成的。但这并不意味着，共同作者们对所讲述的内容都相互完全赞同。而且，有时候某个作者觉得讲义涉及有争议的、远未最终解释清楚的问题。我们不想掩饰这些反对意见和争论，因此“讲课人”从容不迫的讲述有时会被“课堂里的声音”打断，或提出异议。而有时候“课堂里的声音”只是说出经常向我们提到的问题，使“讲课人”的叙述更明确，或加以深化。

为什么书名是《蛋白质物理》？因为我们惊奇地看到一幅图景，生物进化是如何强烈地增强、巩固并向我们证明作为蛋白质中分子相互作用的基础的物理定律的结果。我们还惊奇地看到，我们对基本生物学（尤其是蛋白质系统）的理解是如何通过物理学方法的运用而得以推进。我们可以从蛋白质的质谱法到电子显微镜、到 X 射线和核磁共振研究中观察到这一切。也许，现代科学中没有哪一个其他领域，科学与其哲学的传统边界被如此巨大的益处所冲破。

关于本书中的物理学和生物学。在讲课过程中我们没有放弃讲述物理概念（尤其是统计力学和量子力学的基本原理）的机会。一方面，在我们看来，它们是绝对必不可少的，这不仅是对理解蛋白质的结构和功能而言，而且是对基本科学素养也是如此，“普通的”生物学学生要不就在专业开始时已经完全忘掉了这些物理概念，要不就根本不知道。

另一方面，在有关蛋白质功能的信息海洋中，我们只舀取了对于展示蛋白质空间结构在它们的生物学——更准确些，生物化学——活动中所发挥的作用的绝对必需的内容。

关于实验、物理理论和计算。实验为我们对现象的认识提供了基本事实，以及大量准确的细节。理论让我们理解现象的本质和联系，有助于设计内容丰富的实验。计算把理论和实验联系起来，在理论的关键点进行检验。不过，需要注意的是，并非所有能计算的东西都需要计算出来。例如，蛋白质或水的密度更容易被测量出来。不只是更容易，而且更可信：详细的计算取决于许多很难被精确估算的参数。

在我们的讲义中包括了对一些基本物理理论的讲述，自然，是以最简单的形式。这不仅是因为这些理论让我们能以普适的观点来包容、简化和理解数量众多的实验资料，同样重要的是，这些理论很优美。除此之外，我们认为，对基本的物理

学理论和模型的理解即使对于通识教育也是必不可少的。

关于体内(*in vivo*)**和体外**(*in vitro*)**的实验。**谈到“体内”和“体外”实验,物理学家和生物学家经常有不同的理解。总的来说,并非没有原因:在纯“体内”和纯“体外”之间有许多允许双重解释的层次。例如,在非细胞系统(包含所有核糖体、启动子、终止子,分子伴侣等)中的蛋白质折叠,从物理学家的观点来看很显然是“体内”实验(对他们而说,“体外”应该是溶液中的单个蛋白质;非细胞系统……已经有很多生命实在了!),而在生物学家的观点来看它显然是“体外”实验(对他们来说,“体内”是活的,最好是完整的生物体中)。然而结构学研究,如单个蛋白质,在生物体中实际上是不可能的。因此在实践上总是在合理的妥协中进行的:感兴趣的体内生物学现象用可达到的体外实验观测现象来近似。

关于物理模型,粗略估算和计算机模拟(*in silicio*)①**实验。**在课程进行中我们会经常讨论简单的即与实际相比大大简化的模型,并进行粗略估算。而且我们希望你们在研读完这些讲义之后,也学会做粗略估算,并在你们感兴趣的事件中使用简单模型。初看上去,为什么所有这些是必需的呢?许多人以为,拥有了强大的计算机,可以把“所有的东西”:水分子、盐、蛋白质和 DNA 的原子坐标,以及其他等等,都输入到里面,给定温度,“所有东西都会准确计算出来”。但实际上这是一个乌托邦的图景。计算,我们指的是详细计算(通常是使用所谓“分子动力学”②来进行;对整个蛋白质及其周围环境的量子力学计算基本上不可能),需要花费数天,且只覆盖蛋白质生命中的几个纳秒:因为你必须观察数千个原子的热运动和相互作用。顺便要说到,这种计算不可能达到完全精确:对所有的基本相互作用,我们都只是近似地知道。对系统的描述越详细,考虑到的基本相互作用越多,则在我们的计算中产生的细小误差就越多(更别说所耗时间的增长)。因此所有同样得到的不是“终极真理”,而是多少有点近似的某种估算——超级计算机工作多天的估算。而大家更感兴趣的是简单但快速型的估算:在蛋白质某个位置是否可以引入电荷,或蛋白质在这种情况下是否会爆裂?因此我们的目标之一就是:教会你们做这种粗略估算。

但这并不意味着,我们将会完全忽略计算机实验。这种实验也是实验(通过计算机模拟,而不是体外或体内),而且是用非常复杂的系统:它建立起需要将来解释的事实,从而产生简化的、但清晰的理论。

关于如何处理公式。我们知道,数学公式对生物学家来说是很复杂的,因此尽力避免它们,留下的只是确实必需的。我们想给一个建议:在阅读时请“用公式来

---

①*in silicio*(以前曾写为“*in silico*”)表示“在计算机上进行的”,或“通过计算机模拟的”之意,有时候也翻译为“硅片实验”。它只应用于模拟自然科学中的自然过程或实验室过程,而不表示由计算机执行的通用计算。

②在分子动力学中,使用经典力学(而非量子力学)对原子进行描述。

检验文字”！当然，只阅读文字更简单，但是它们常常会产生歧义；因此，用公式检验文字并反向进行，你们可以检查自己的理解程度。

关于小字体文字。它们给出了有益的，但并非必需的题外话，补充和注释。

关于口味。这些讲稿，毫无疑问，表现了我们的个人口味和嗜好。它们更专注于理解现象的本质，而非细致描述它们的细节；在讲义中有很多理论和物理问题，仅有最少的必需的实验事实（几乎没有实验技术；尤其是我们实际上根本没有讲述X射线晶体学和NMR谱学，这些方法是我们为关于蛋白质结构的知识而最应该感谢的）。因此我们的讲义决不能代替正规的富含事实的生物学和生物化学“蛋白质”课程，也不应该作为手册来使用。涉及具体的蛋白质，我们只给出了最重要的（从我们的观点出发！）例子；表格中只包含绝对必需的数据；图片特意做成概要性的；所有数字舍去了尾数；如此等等。

关于这些讲义中的个人成分。我们没有放过机会提到我们自己和我们在俄罗斯科学院蛋白质研究所的同事在蛋白质物理方面所做的工作。这当然加大了讲义的“个人成分”，而我们希望能使它们更有生气……

目　录

蛋白质的空间结构

绪　论

第1讲

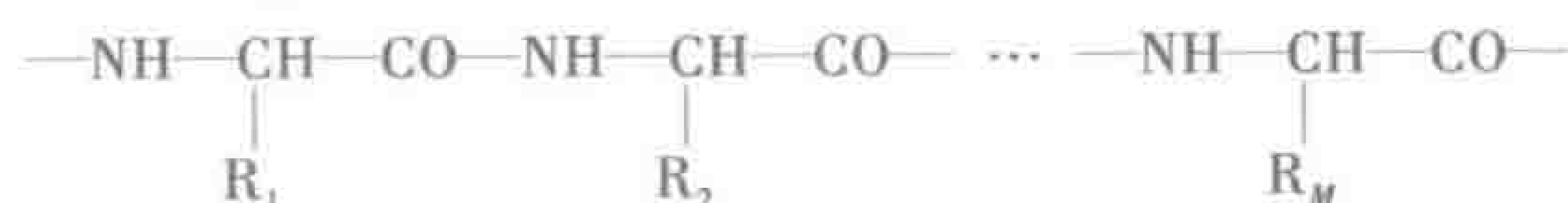
蛋白质的主要功能。氨基酸序列决定空间结构,空间结构决定功能。反之不然。球状蛋白,纤维蛋白和膜蛋白。蛋白质的一级、二级、三级和四级结构。蛋白质的生物合成。蛋白质在体内(*in vivo*)和在体外(*in vitro*)时的折叠。翻译后修饰。

蛋白质是分子机器、建筑模块和活细胞的铠甲。蛋白质最重要的,也几乎是唯一的功能,是对细胞内和细胞周围的化学转换进行酶促催化作用。除此之外,调节蛋白控制基因表达,受体蛋白(它们位于脂膜上)保障细胞间信号的感知,而信号常常是由激素(也是蛋白质)传递的。免疫蛋白和与之类似的组织相容性蛋白质既能识别和结合"异种"分子,也能识别和结合"自身的"细胞,从而让这些细胞能正确地安置在生物体中。结构性蛋白构成微丝和微管,以及纤丝、毛发、丝和其他保护性覆盖物,它们加固细胞膜,支撑细胞和组织的结构。运输蛋白转运(而储存蛋白——储存)其他分子。跨膜传递质子和电子的蛋白质保障所有的生物能学过程:光吸收,呼吸,ATP 生产。其他蛋白质,通过"燃烧"ATP,保障机械化学活动,它们在肌肉中工作或移动细胞成分。

尽管有这些多样性,蛋白质的工作总是以它们和与之发生作用的分子之间的相互作用的高度专一性:就像钥匙和锁(更准确些:像柔软的钥匙和柔软的锁)为基础的。为了保障这种相互作用,蛋白质必须有足够"坚硬"("工作中的"蛋白质在任何情况下)的空间结构。所以蛋白质(以及对生命非常重要的其他大分子:DNA 和 RNA)的生物学功能与它们特定的三维结构密切相关。这些结构的破坏,甚至只是小小的改变,常常都会导致蛋白质活性的丧失或显著改变。

有关蛋白质分子三维结构的知识对理解蛋白质分子的功能是必不可少的。因此我将在最后才讲述蛋白质功能的物理学,开始时集中注意力于它们的结构,关注它们的稳定性和自组织能力的本质。

■ 蛋白质是高分子聚合物①:由氨基酸残基组成的多肽链②;这是由 E. Fischer 于 20 世纪初所发现的。在 20 世纪 50 年代早期,Sanger 证明蛋白质链具有唯一的链节序列,其链节是氨基酸残基。“残基”是游离氨基酸聚合成为蛋白质链之后留下来的部分。该链具有化学上规则的骨架(“主链”),从骨架上伸出各种不同的氨基酸侧链基团:R_1 原子团,R_2 原子团,……,R_M 原子团。



蛋白质链中的链节数目 M 是由基因编码的,有时是几十个,有时可达几千个。

存在有 20 种基本氨基酸残基。它们在蛋白质链中的位置是由基因编码的。但是后续的蛋白质修饰有时会增加氨基酸的多样性。此外,在某些蛋白质中还包含其他分子(辅因子)。

在“工作中的”蛋白质中,肽链以严格确定的方式折叠起来。在 20 世纪 50 年代末期,Perutz 和 Kendrew 解析出了第一批蛋白质空间结构,并证明了它们的结构具有高度复杂性和唯一性。然而,蛋白质分子结构的严格确定性实际上是在 19 世纪 60 年代由 Hoppe-Zeiler 首次证明的(这是后来才弄清楚的),他获得了血红蛋白的晶体:因为在晶体中每个分子的每个原子都占据唯一的位置。

蛋白质在晶体中和溶液中的结构是否相同?这一问题讨论了多年:当时只有间接的数据可以证明,直到最后被核磁共振(nuclear magnetic resonance,NMR)证实,实际上通常是一致的(在误差范围内)。

蛋白质“生活”在各种不同的环境中,其结构带有环境的明显烙印。蛋白质周围的水越少,越难弥补那些拉紧肽链的强力氢键的断裂,这些氢键(正是它们稳固了蛋白质分子的骨架结构)对蛋白质就越重要,蛋白质的稳定结构就必须越规则。

粗略地说,根据“生活环境”和基本构造类型,蛋白质可以分为三类。

1. 纤维蛋白:它们形成巨大的聚集体,其中极少有水;它们的结构通常非常规则,基本上由不同的链之间的相互作用所维系。

2. 膜蛋白:它们“生活”在没有水的膜中,但某些部分从膜内伸出到水中。这类蛋白质的膜内部分(纤维蛋白也是如此)非常规则且由氢键结合,但是这些规则部分的尺寸受限于膜的厚度。

3. 水溶性球状蛋白:它们生活在水中,最不规则(尤其是较小的);其结构由蛋

①高分子聚合物(polymer,也常简称为“高分子”)指由许多相同的、简单的结构单元(或称为链节、单体)通过聚合反应后由共价键重复连接而成的相对分子质量很高(通常可达 $10 \sim 10^6$)的化合物。如果所有结构单元是完全相同的,则称为同质聚合物(同聚物),如果聚合物包含多种不同的结构单元,则称为异质聚合物(异聚物、杂聚物或共聚物)。橡胶、纤维和塑料等都是常见的高分子聚合物,而蛋白质、DNA 和 RNA 等生物分子也都是高分子聚合物。

②由两个氨基酸分子脱水聚合而成的化合物称为二肽,同理类推还有三肽、四肽、五肽等。通常称由 20 个以上的氨基酸残基组成的化合物称为多肽。(也有文献把由 2~10 个氨基酸组成的肽称为寡肽或小分子肽;10~50 个氨基酸组成的肽称为多肽;由 50 个以上的氨基酸组成的肽就称为蛋白质。)

白质链自身的相互作用所维系，而且链上相距较远、但空间上接近的烃基[①]("疏水的"，即"怕水的")基团之间的相互作用尤其重要，有时候蛋白质链与辅因子之间的相互作用也很重要。

最后，较小的，或者缺少烃基基团的多肽可能自身并不具有固定的结构，但是在与其他大分子相互作用时而获得固定结构。

无疑，上述分类是极为粗糙的。有时候蛋白质可以由纤维状的"尾部"和球状的"头部"(如肌球蛋白)构成，如此等等。

迄今(2012年)已经知道几百万条蛋白质氨基酸序列(已经建立了专门的计算机数据库储存它们，如UniProtKB[②])和约十万个蛋白质空间结构(它们储存在计算机蛋白质结构数据库中：Protein Data Bank，缩写为PDB[③])。已知三维结构的蛋白质，绝大部分都属于水溶性球状蛋白。对于膜蛋白和纤维蛋白，已解析出的仅仅是相对较少的空间结构或它们的独立片段。原因很简单：水溶性蛋白容易分离成单个分子，在溶液中它们的结构容易用X射线研究(在晶体中，或者用核磁共振谱研究)。因此，讲到"蛋白质结构"、"蛋白质结构的形成"等时，实际上指的是仅仅在水溶性球状蛋白中已证实的规律。在阅读书籍或文章时，应该对此牢记在心，阅读本书也是如此。此外，还需要记住，由于实验的原因，当代所有的蛋白质物理学，也主要是关于小蛋白质的。大蛋白质分子及其复合体的物理学，才刚刚开始发展。

■　维系蛋白质空间构造的非共价键相互作用，比在蛋白质链中固定单体(残基)序列的共价键要弱得多。这种氨基酸残基序列被称为"蛋白质一级结构"(图1-1)这是在根据写在基因中的"指令"而进行的生物化学基质合成过程中生成的。

蛋白质的架构(archetecture)[④]，尤其是球状的水溶性蛋白，非常复杂且种类繁多，与DNA双链单一性全然不同(顺便说一句，单链RNA像是介于两者之间)。但是在蛋白质中发现了一套"标准"的结构，关于它们我们将会非常详细地讲到。

在这里首先要谈到蛋白质的规则二级结构：α螺旋和β结构；α螺旋通常用螺旋状的丝带(图1-1)或圆筒表示，而伸直的β结构片段(它们黏在一起会形成薄片)用箭头表示(图1-1)。二级结构的特点是主链具有规则的周期性形状[或者如通常说的"构象(conformation)"]，尽管侧链基团的构象各不相同。

---

①碳原子和氢原子所构成的化合物称为烃类。烃类分子中去掉一个或几个氢原子而成的基团称为烃基。例如，甲基 CH_3—、乙基 CH_3CH_2—、苯基 C_6H_5—等。

②UniProt(http://www.uniprot.org/)是信息最丰富、资源最广的蛋白质数据库。它由整合Swiss-Prot、TrEMBL和PIR-PSD三大数据库的数据而成。UniProtKB(UniProt知识库)主要由两部分组成：UniProtKB/Swiss-Prot(包含手工注释、已校验的条目)和UniProtKB/TrEMBL(包含自动注释、未校验的条目)。

③蛋白质结构数据库(PDB)是美国纽约Brookhaven国家实验室于1971年创建的。1998年由美国国家科学基金委员会、能源部和卫生研究院资助，成立了结构生物学合作研究协会(RCSB)来管理PDB(http://www.rcsb.org/pdb)。

④本书中"archetecture"翻译为"架构"，指的是骨架结构。



图 1-1 蛋白质结构的组织水平：一级结构（氨基酸序列）；二级结构（图中显示了 α 螺旋和一串 β 结构）；单链折叠成的三级球状结构；由几条蛋白质链形成的寡聚蛋白质的四级结构（此处为二聚的 cro 阻遏蛋白①）

一条多肽链的二级结构折叠成的球体，称为三级结构（图 1-1）。几个多肽链组合而成的“超球”，称为蛋白质的四级结构（图 1-1）。例如（作为显示于图 1-1 中的 cro 阻遏蛋白二聚体的补充）：血红蛋白由两条 β 链和两条 α 链②（与 α 螺旋和 β 结构没有任何关系！）组成。如果蛋白质的四级结构是由同样的链所组合而成，则它通常会是对称性的（cro 阻遏蛋白和血红蛋白在此都不例外）。有时四级结构会由几十条链组合而成。特别是病毒外壳蛋白可以看做是这种“超四级”结构。

在三级结构中有一些结构是极为典型的，我们以后将会了解它们。这些典型结构常常并不包含整个球状蛋白，而只是包含它的紧密亚球：“结构域（domain）”（或简称为“域”）（图 1-2）。结构域（与小球状蛋白类似）通常由 100～200 个氨基酸残基，即约 2000 个原子组成。其直径为 30～40Å。

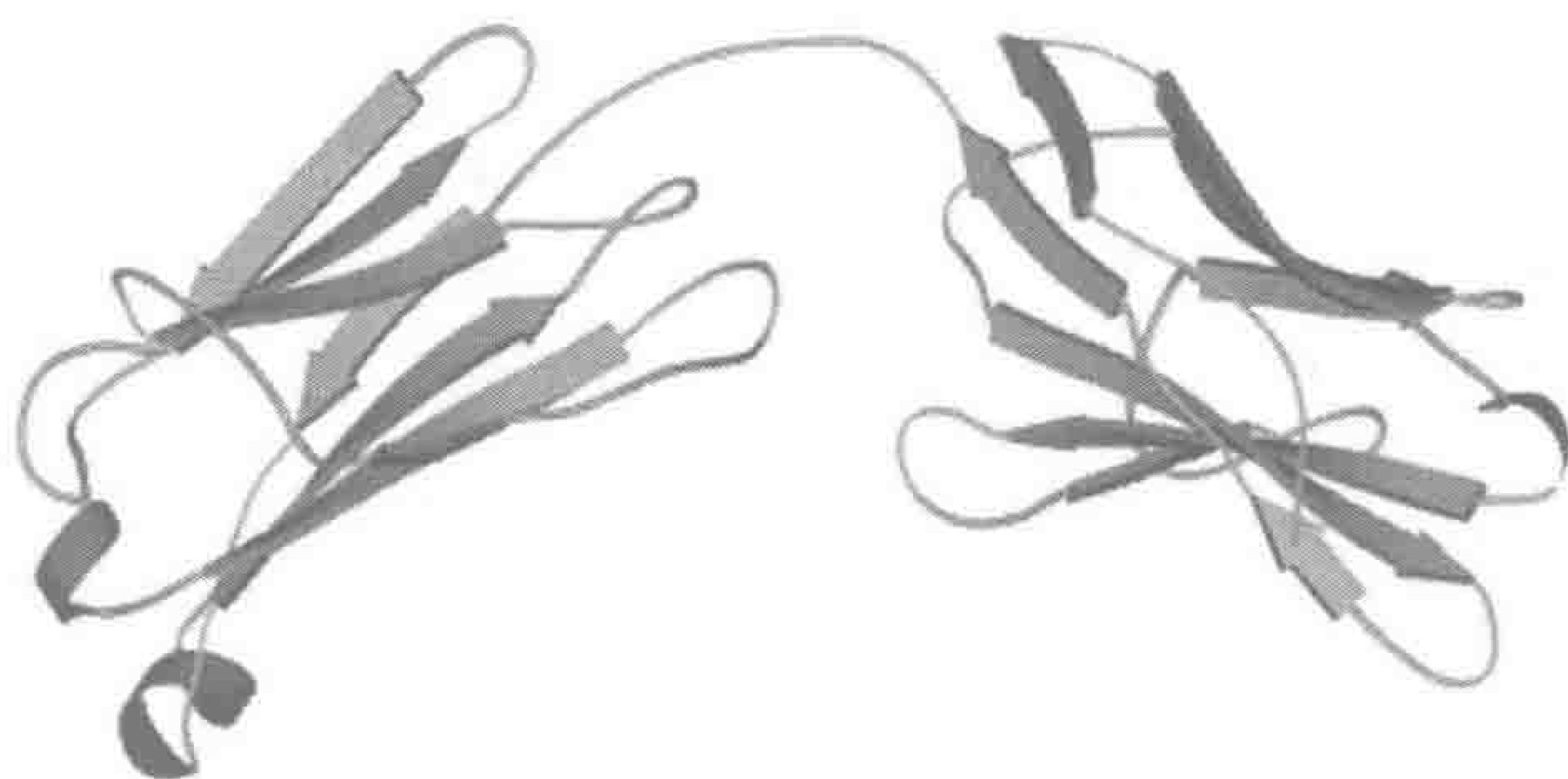


图 1-2 大蛋白质的域结构类似于小蛋白质折叠起来的四级结构。差别仅仅在于，大蛋白质的所有紧密亚球在域中是由单链构成的，而四级结构是由多条链组合而成

■ 在体内，蛋白质三维结构的形成发生于生物合成过程中或紧随其后。然而，值得指出的是，它不仅仅可以发生在生物合成中：Anfinsen 于 1960 年左右证实，

①阻遏蛋白：结合在操纵基因上的一种蛋白，阻止操纵子以及操纵子的酶的转录。

②该分子包含 4 条链，残基序列是两两相同的，这两种链按希腊字母编号为“α”和“β”。

它可以发生于“复性”过程中——在体外情况下由于某种原因而解折叠的蛋白质链重新折叠,而且完全是自发进行的。

这意味着,氨基酸序列本身(在合适的温度和水 pH 下!)决定了蛋白质的空间结构,即蛋白质能够自组织①(self-organizing)。

课堂里的声音: 严格地说,这里指的主要是相对较小(最多 200~300 个氨基酸残基)的水溶性球状蛋白。更大的球蛋白,尤其是高等生物的蛋白质,情况更为复杂:它们远不是全部都能自发地复性……

讲课人: 谢谢你的准确化。确实,那样的蛋白质更为复杂。部分原因是聚集作用,部分是翻译后修饰,尤其是在真核细胞中。对膜蛋白和纤维蛋白的自发性自动折叠知道得仍然比较少。在一些那样的蛋白质中确实发生,但通常无法达到复性。因此应当在此预先声明:讲述蛋白质物理学、蛋白质的结构及其形成时,我实际上(如果未作特别说明)通常指的相对较小的球蛋白。这是在生物物理学文献中通常的预先约定,但是未清楚地说明,经常导致误解。

■ Anfinsen 的实验原则上把一个物理过程(蛋白质空间结构的自组织)与蛋白质链的生物化学合成过程区分开来了。现已清楚,蛋白质的结构完全由它的氨基酸序列所决定,蛋白质不受细胞机器的影响。看来,细胞机器的主要任务是保护正在折叠的蛋白质不受不期望的接触(其中包括极大蛋白质中距离较远、形成不同结构域的链块之间的接触)的影响,因为在体内蛋白质是在细胞质中进行折叠,其中有那么多的不同分子、那么多黏附到某些东西上的可能性!但是在体外的稀溶液中,蛋白质(至少是小蛋白质)能完全独自自发地折叠起来。

■ 严格地说,蛋白质要能自发地自动折叠和复性,仅仅只有它不经受强烈的翻译后修饰才行,即它的化学结构在生物合成和初始自动折叠后不被剧烈破坏。例如,胰岛素(insulin)(它的一半链在折叠后被丢弃)就不能复性。

翻译后修饰花样繁多。通常,化学修饰是由特定的酶控制的,而不是蛋白质自身的“自组织”进行的。除了蛋白质链的断裂[蛋白质水解:它经常能够把酶原(未活化的“蛋白质前体”)转变为活性酶;同时,它也能把多聚蛋白分解成多个独立的蛋白质球体]之外,还观察到有蛋白质链的小块切除(例如,由胰岛素原制作为胰岛素时。顺便说一句,切下的片段有时候就成为激素)。还观察到有链末端的修饰,乙酰化,糖基化,脂质结合到链上特定位点,某些特定的侧链基团的磷酸化,诸如此类。最近甚至观察到蛋白质链的“剪接(splicing)”(自发地切除链块并在把切开的两端黏合起来)。偶尔也观察到蛋白质链或它的某些小块

---

①在不存在外部指令的情况下,系统按照相互默契的某种规则,各尽其责而又协调地自动地形成有序结构,就是自组织。对于蛋白质而言,“自组织”主要是指“自动折叠”。

自发的环化。

特别需要注意在含硫的半胱氨酸(Cys)残基之间的二硫键(S—S)的形成:尽管在体内它们的形成是由专门的酶(二硫键异构酶)催化的,但是在合适的体外条件下能够自发地自动折叠形成"正确的"二硫键。通常,这些键只在分泌型蛋白质中结合生成(细胞内没有氧,因此没有合适的氧化势);因此如果这些二硫键能够"正确地"被结合,决不会妨碍、而是更有利于促成蛋白质复性。

相反,"不正确地"结合的二硫键会妨碍蛋白质的复性能力。顺便说一句,正是如此,使煮熟的鸡蛋不能在冷却时"还原"。原因在于,在加温时鸡蛋的蛋白质不仅变性,而且在相互之间以额外的二硫键交联(就像在橡胶中那样)。在冷却时这些化学键仍然保留着,不允许鸡蛋的蛋白质恢复到初始("天然")状态。

■ 总之,氨基酸序列决定了蛋白质的空间结构,而这种结构决定了它的功能,即这个蛋白质与谁发生相互作用、干什么。

■ 在这里有必要作几点注解。

1. 观看图 1-1 时,有人可能会想,蛋白质内部有很多空余空间,因此,它是"软的"。但实际上完全不是这样!蛋白质是硬的:蛋白质链紧凑地拼装起来,原子紧挨原子(图 1-3 左边部分)。那个图中显示的只是蛋白质的解剖结构,它的骨架。它的内部结构在全原子模型中不方便研究:只有让原子"透明"、把注意力集中到蛋白质链的通路(见图 1-3 右边部分,特别地,见图 1-1 和图 1-2,其中侧链基团被"剥落",突出显示二级结构元件),才能看见蛋白质的内部构造。在蛋白质全原子模型(图 1-3)中,甚至看不出蛋白质是由聚合链形成的:只能看到球体表面,像个土豆。不过在研究蛋白质功能时需要用全原子模型[它的功能专一性正是由球体(这个"土豆")表面的物理-化学和几何性质所决定的,而蛋白质骨架]建立并固定这个表面。

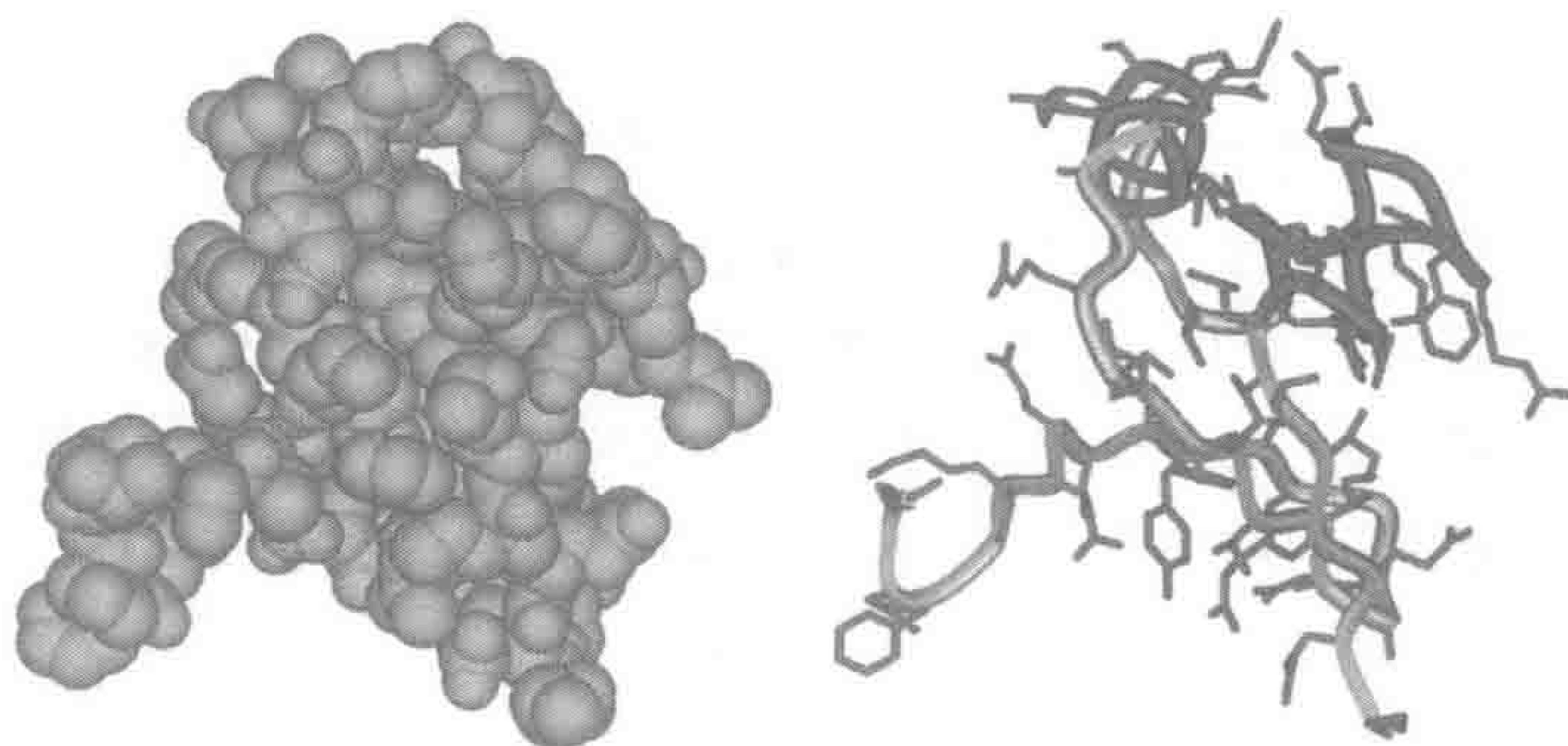


图 1-3 蛋白质球的原子模型(左)与其骨架模型(右)。在骨架模型中侧链基团显示为细线条,主链显示为粗细条

2. 在蛋白质组分中,除了蛋白质链本身之外,常常还有辅因子(图 1-4):小分子、

离子、糖、核苷酸和核酸片段等。这些非肽类分子参与行使蛋白质功能，有时还参与蛋白质分子结构的形成。辅因子与蛋白质可能以共价键连接，也可能仅仅是附着于蛋白质球体的特定位置。此外，通常有很多水分子紧密地附着在蛋白质表面。

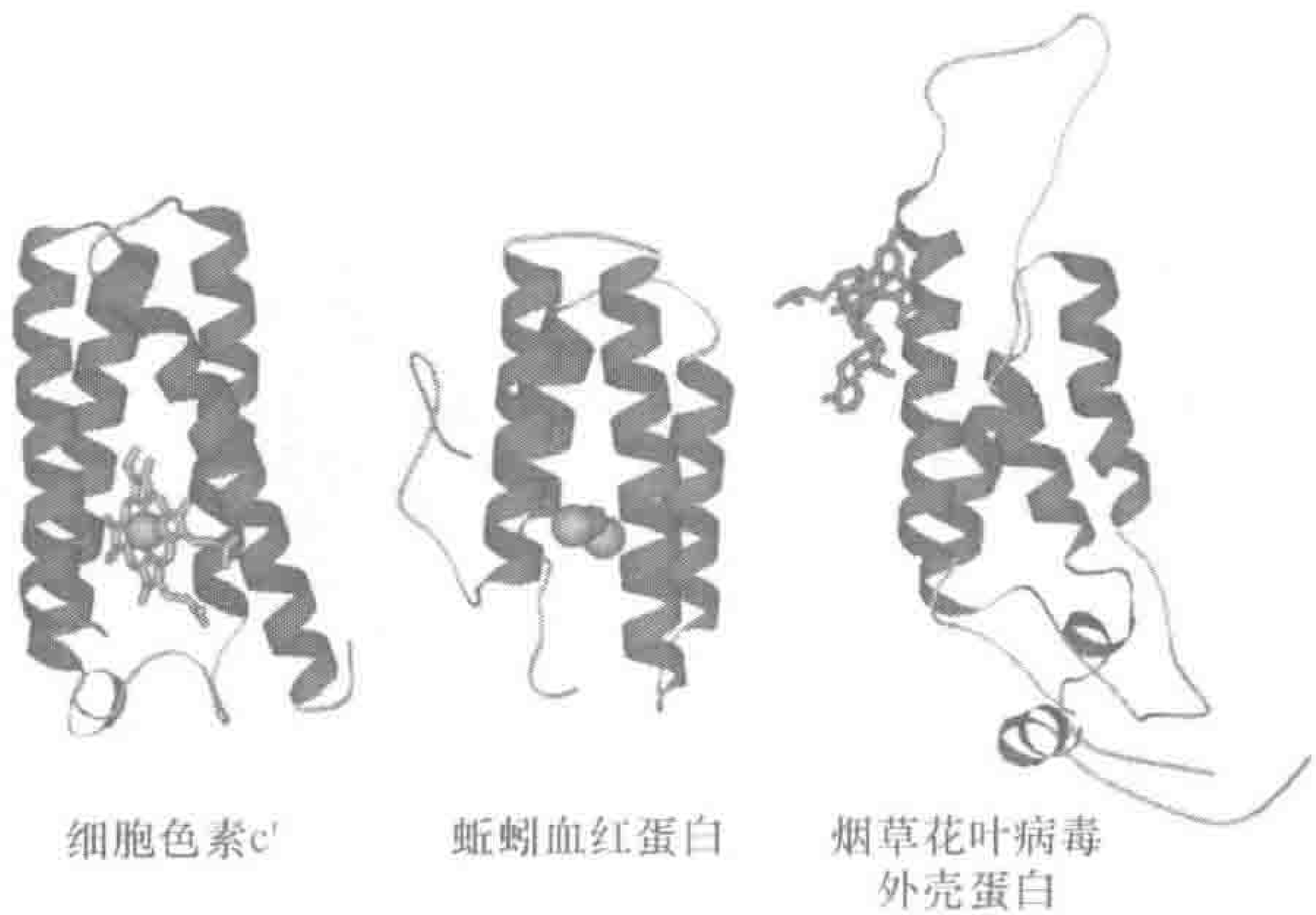


图 1-4　3 个形状相似(每个都是由 4 个 α 螺旋组成)、但功能不同的蛋白质:细胞色素 c′、蚯蚓血红蛋白和烟草花叶病毒外壳蛋白。图中显示了蛋白质链和辅因子:骨架模型[亚铁血红素(细胞色素中)和 RNA 片段(病毒外壳蛋白中)];小球[铁离子(细胞色素和蚯蚓血红蛋白中)和与铁结合的氧(蚯蚓血红蛋白中)]

3. 在外部条件变化(如温度升高)时，坚硬的蛋白质(薛定谔①所说的“非周期性晶体”)表现得就像晶体，即它能“忍耐”到某种程度，然后一下子熔化，而不是像玻璃那样逐渐失去其形状和硬度。蛋白质的这种基本性质与其工作的可靠性是紧密相关的:像电灯泡一样，蛋白质的破裂是遵循“全或无”原则的，而不是渐变的(否则会导致它的行为不可靠，如失去专一性等;对此我们还会详细讨论)。

最后，讲到硬度时，需要区别相对较小且确实坚硬的单域蛋白质(它们由一个紧密的小球构成)和大得多的包含多个域(图 1-2)或具有四级结构(图 1-1)的蛋白质:构成它们的亚球可以移动或散开。

当然，除此之外，在蛋白质行使其功能时，所有的小球都像固体一样，可以轻微地变形(但不是完全变形!)

■　通常，内部结构(“解剖结构”)相似的蛋白质行使相同的功能。例如，许多(尽管

---

①E. 薛定谔(Erwin Schrdinger,1887～1961)，奥地利理论物理学家，波动量子力学的创始人。他后来转向生命科学研究，并于 1944 年发表了《生命是什么?——活细胞的物理面貌》一书。在书中，他试图用热力学、量子力学等理论来解释生命现象和本质，引入了负熵、遗传密码、量子跃迁式的突变等概念，这些概念至今仍有着广泛的影响。这本书使许多青年物理学家开始注意生命科学中提出的问题，引导人们用物理学和化学方法去研究生命的本性，使薛定谔被公认为是当今蓬勃发展的分子生物学的先驱。书中用量子力学的观点论证基因的稳定性和突变发生的可能性，其中提出必定有一种由同分异构的连续体构成“非周期性晶体”，其中含有巨大数量的排列组合，编排成遗传密码。他最初推测是蛋白质(但现在已证明是 DNA)。

并非全部)细胞色素都如图 1-4 中显示的那样;而许多丝氨酸蛋白酶①(尽管同样并非全部)(无论是来自于脊椎动物或来自于细菌)都与胰凝乳蛋白酶很相似(图 1-5)。

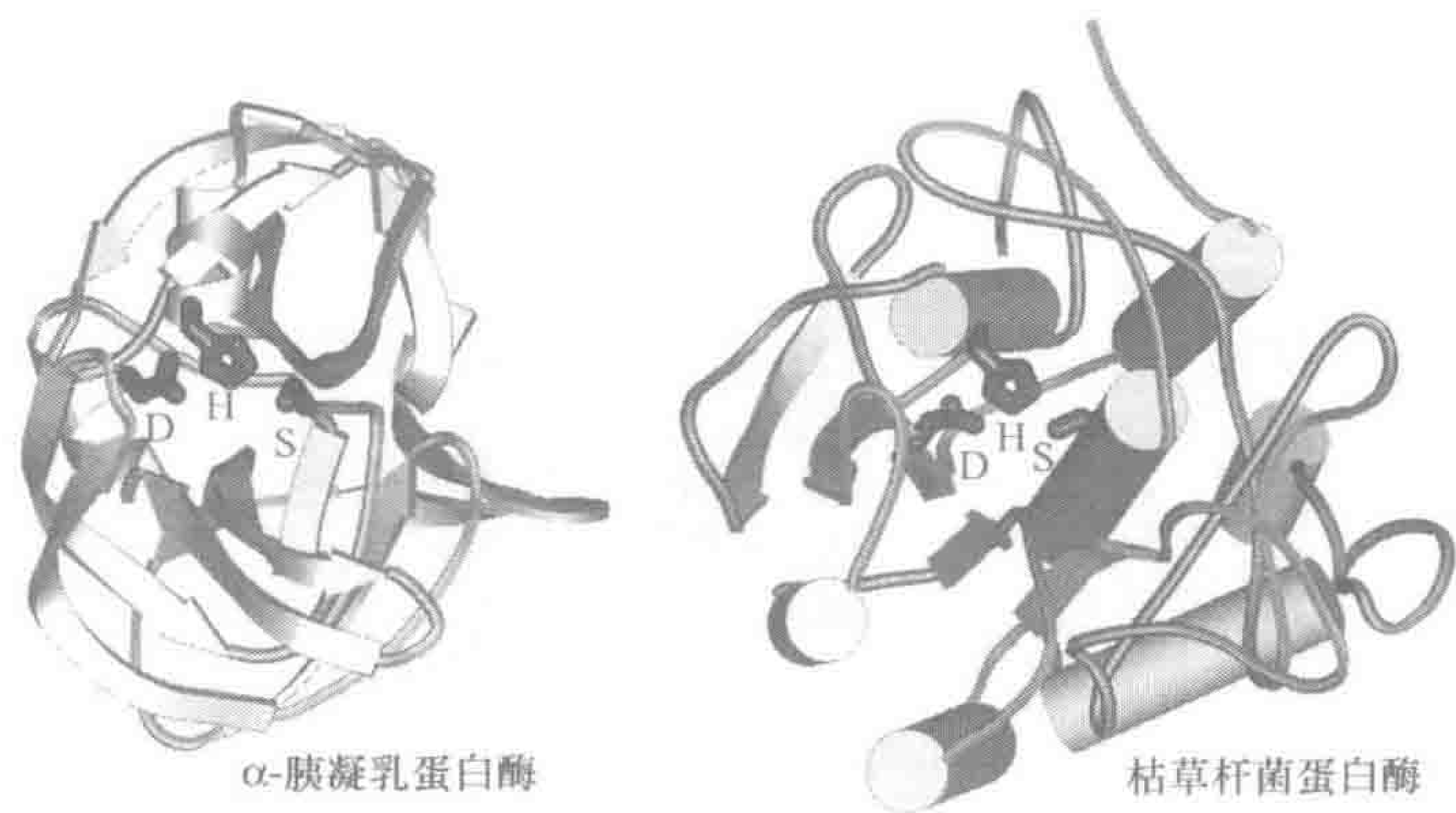


图 1-5　两个形状不同,但功能几乎完全一样的蛋白质(丝氨酸蛋白酶②):由 β 结构组装成的胰凝乳蛋白酶和除了 β 结构之外还包含多个 α 螺旋(其中几个进入了活性中心)的枯草杆菌蛋白酶。尽管肽链的折叠很不相同,但是它们的催化中心由相同的且在空间排列位置(不是在肽链中的排列位置!)也相同的残基所组成:胰凝乳蛋白酶中是 Ser195 (S)、His57 (H) 和 Asp102 (D),枯草杆菌蛋白酶中是 Ser221 (S)、His64 (H) 和 Asp32 (D)

然而有时候实际上相同的空间结构可能行使完全不同的功能。例如,在图 1-4 中所显示的空间结构大致相似的三个蛋白质,一个(细胞色素)与电子结合,另一个(蚯蚓血红蛋白)与氧结合(这两个功能有点相似,两者都包含于氧化反应链中),第三个(病毒外壳蛋白)与更大得多的分子 RNA 相结合,其他外壳蛋白也是如此。

■　前面已经讲到,蛋白质的结构决定它的功能。

反过来是否成立?即蛋白质的功能是否决定它的结构呢?虽然找到了部分的这种相关性,但是总的来说,功能的这种确定性影响只存在于非常粗略的水平,由蛋白质功能的"环境条件"所决定的水平。例如,承担结构性功能的蛋白质(如组成毛发或纤毛的)通常是纤维蛋白;受体通常是膜蛋白,如此等等。但是在蛋白质

①丝氨酸蛋白酶是一个蛋白酶家族,其作用是断裂大分子蛋白质中的肽键,使之成为小分子蛋白质。其激活是通过活性中心中的一组氨基酸残基的变化实现的,其中一定有一个是丝氨酸(名称来源于此)。在哺乳类动物中,丝氨酸蛋白酶扮演着很重要的角色,特别是在消化、凝血和补体系统方面。在消化酶中,胰分泌的酶里面有三种是丝氨酸蛋白酶:胰凝乳蛋白酶(糜蛋白酶)、胰蛋白酶和弹性蛋白酶。这三种酶的一级结构和三级结构都相似。

②蛋白酶:催化蛋白质水解的酶。胰凝乳蛋白酶是一种胰消化酶,在小肠中催化某些蛋白质水解形成多肽和氨基酸。枯草杆菌蛋白酶是由一种土壤细菌产生的细胞外蛋白酶,能够将蛋白质分解成多肽,其作用过程类似胰蛋白酶。

架构水平，即它的内部构造水平，这常常完全没有这种相关性。例如，两种丝氨酸蛋白酶（胰凝乳蛋白酶和枯草杆菌蛋白酶）具有相同的催化功能和几乎相同的专一性，但是内部结构完全不同（图 1-5）（这些蛋白质的相似性并不比海豹与龙虱的相似性更大：只有它们的“鳍”，（即它们的活性中心，由肽链中几百个链节之中的六个残基组成）的结构是完全一模一样的，而其他部分完全不同）。此外，存在着行使同样功能（例如，丝氨酸蛋白酶和金属蛋白酶的）的不同结构的活性中心。

总之，一系列例子表明，蛋白质的功能不决定其空间结构。但这样说的时候，必须注意尺度。如果与蛋白质发生作用的分子很大，几乎整个蛋白质与之发生作用，则蛋白质的整体形状对其功能是很重要的。如果与蛋白质发生作用的分子较小（这更常见），则仅由蛋白质表面上较小的部分的小尺度细节决定它的功能，而其余部分仅仅是支撑这些关键的重要细节。后者意味着蛋白质链的大部分的主要任务是为活性中心建造一个坚硬的底座。或者说，蛋白质功能的专一性意味着它的坚硬性。

课堂里的声音：架构完全不同的蛋白质可以完成同样的功能，而不同的功能可以由结构相似的蛋白质执行，强调这些非平凡而令人兴奋的事实的时候，不应该忘记陈旧的、但同样可靠的原则，即架构相似的蛋白质经常是同源的（遗传相关），并且具有相同或相似的功能……

讲课人：确实如此，但没什么意义。而我想强调对于蛋白质物理学非常重要的观点，即活性中心可能很少依赖于蛋白质整个其余部分的构造。

“整个其余”部分的共同性质是，在任何条件下“工作的”蛋白质的身体是坚硬的：否则无法保证活性中心的活动专一性。

■ 在接下来的时间里我们将详细地考虑蛋白质的结构、自组织能力和坚硬性的原因，将讨论蛋白质的功能和生物学家感兴趣的其他问题，但开始时必须了解氨基酸残基，了解它们内部的或与周围环境的基本相互作用。

蛋白质内部和周围环境的基本相互作用

第 2 讲

L 型氨基酸残基的立体化学。共价键连接和它们之间的角度。共价键的振动。围绕共价键的转动。肽基团。顺式(*trans*)和反式(*cis*)脯氨酸(Pro)。

构成多肽链的氨基酸(图 2-1)有两种立体构型:L 型和 D 型(参见彩色立体图对 D-1 和 D-2)。

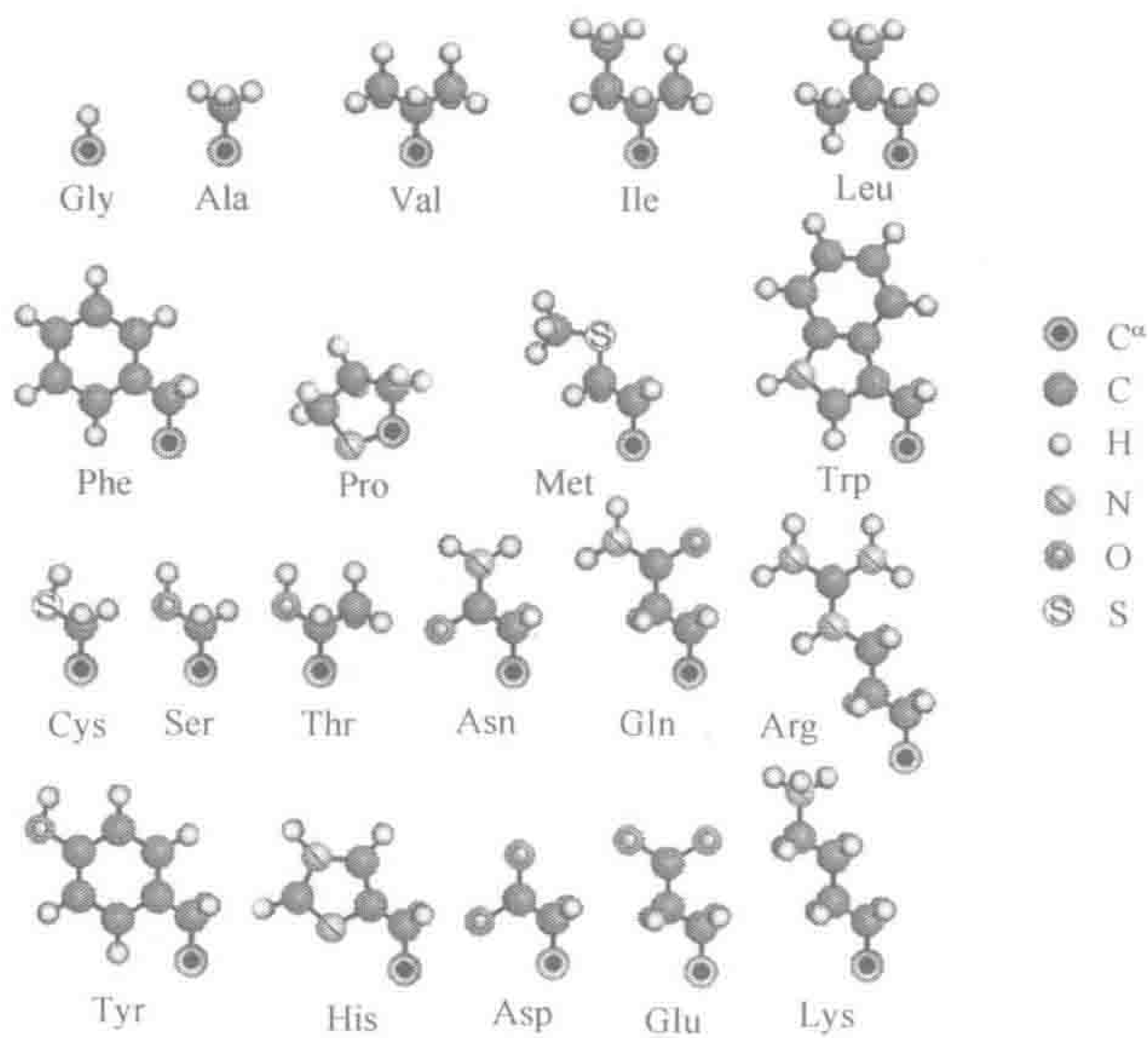
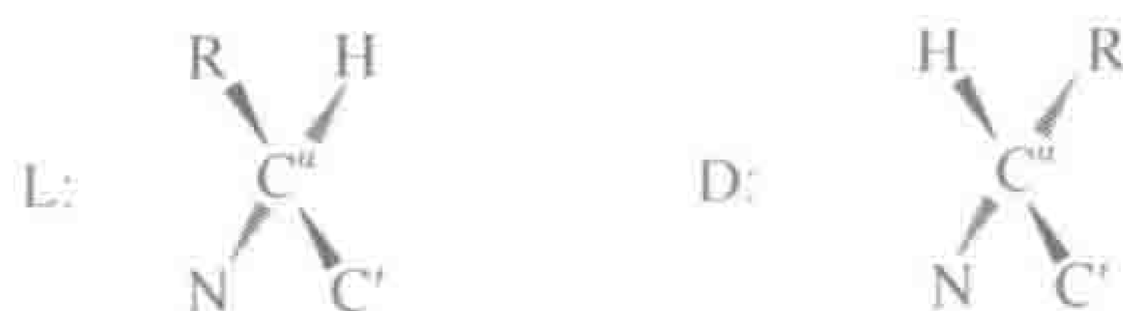


图 2-1　20 种标准氨基酸残基的侧链,它们从主链上的 C^{α} 原子向外伸出。图例为氨基酸所含原子

L 和 D 这两种构型是镜像对称的:与 α 碳原子(C^{α})相连的重侧链基团 R(显示于图 2-1 中)和 H 原子,在它们中互换位置(图中箭头与符号的大小和亮度减小指示原子的凹陷处):



仅仅只有甘氨酸(Gly)没有 L 和 D 型,因为它的“侧链”仅由 H 原子组成。

蛋白质链仅由 L 型氨基酸残基构成。只有这些残基是由基因所编码的。D 型氨基酸残基(它们存在于多肽中)没有被编码在蛋白质的基质合成中,而是由特定的酶所合成的。自发的外消旋(L↔D 转变)现象在蛋白质中实际上并不发生。它在生物合成中也不发生,但是经常在肽的纯化学合成中观察到,并且很难消除。

蛋白质链中的氨基酸是由 C′ 和 N 原子之间的肽键相连接的(图 2-2)。

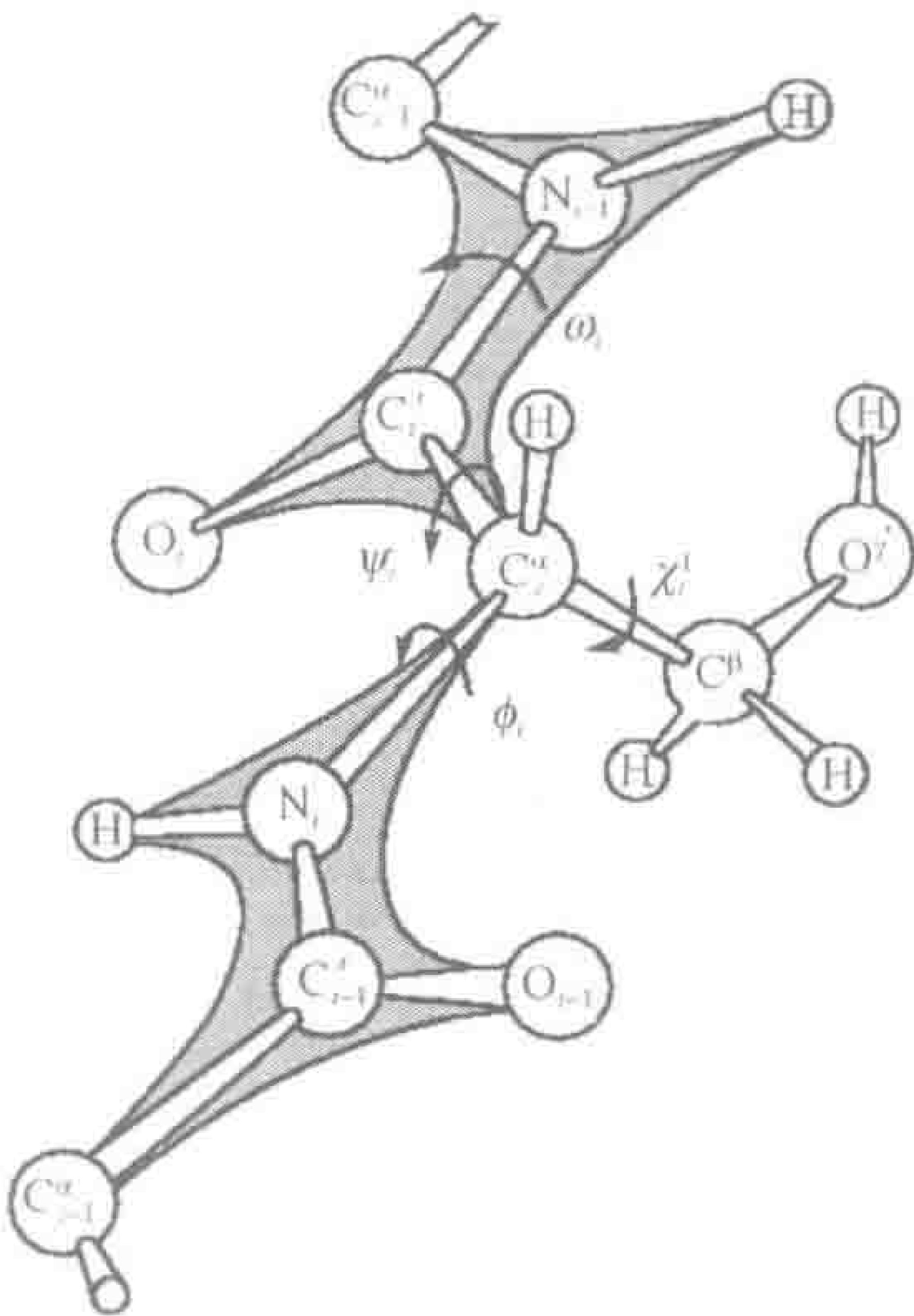
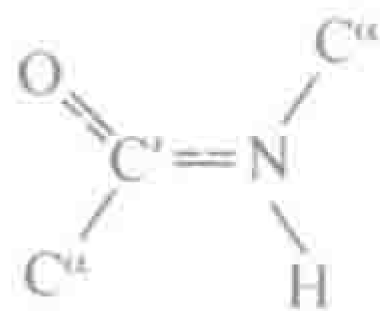


图 2-2　多肽:主链和与之相连的侧链(Ser)。肽基团以阴影表示。图中显示了主链上的内旋转角(ϕ,ψ,ω)和侧链上的内旋转角(χ^1)。下标 i-1,i 和 i+1 为链上的氨基酸残基序号。箭头表示靠近读者的链端相对于远离读者的链端旋转时角度增大的方向。图片取自附录立体图对 D-3,有改动

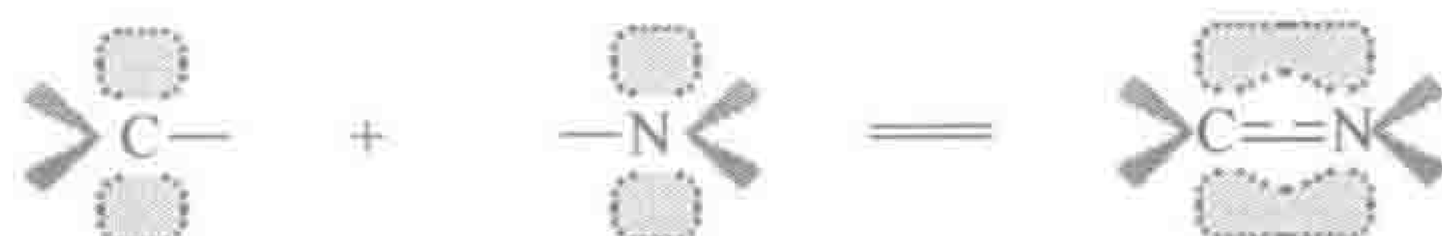
■　所有肽单元的平面构形和刚性在蛋白质结构中起着重要的作用。



肽单元的平面构形是由 N 和 C′原子中电子的所谓 sp^2 杂化作用所赋予的。电子轨道的“杂化”完全是量子效应。sp^2 杂化把原子的一个球形 s 电子轨道和两个“8 字形”的 p 轨道改变成为三个从核向外伸展的 sp^2 轨道。这三个轨道使该原子进入到处于同一平面内(—●)的三个共价键之中。共价键是由围绕两个成键原子的“离域”(delocalized)电子云所形成的。

肽单元的刚性来源于未参与 sp^2 轨道的 p 电子的额外连接。N、C′ 和 O 原子的这些 p 电子结合成键并“离域化”,从而产生包围 N、C′ 和 O 原子的电子云(正因为如此,C′== N 和 C == O 被绘制成相当“一个半”的==形状)。而且由于 p 轨道具

有垂直于所有三个 sp^2 轨道(—●)的"8 字形",所以这些"垂直的"p 轨道的额外共价键能阻碍绕 C′—N 键的转动。



我想要提醒一下,电子的离域化,从一个原子转移到另一个原子,是形成共价键的主要原因。这可从海森堡不确定性原理推导出来。

$$\Delta p \Delta x \sim \hbar \tag{2.1}$$

式中,Δp 为粒子动量的不确定度;Δx 为坐标的不确定度;而$\hbar \equiv h/2\pi$,其中 h 为普朗克常量。由于电子在原子中的运动方向无法知道,故 $\Delta p \approx |p| = m\nu$,其中 ν 为粒子速度,m 为粒子质量。因此,

$$\nu \sim \hbar/(m\Delta x)$$

同时,粒子的动能为 $E = m\nu^2/2$,则

$$E \sim \hbar^2/(m\Delta x^2) \tag{2.2}$$

因此,在离域化时,粒子的能量随 Δx 的增加而减小,粒子进入更稳定的状态。可见,这种效果对轻粒子(即电子)更加强烈。这样,电子的离域化导致产生共价键。

共价键的键长接近于原子的范德华半径,即 1~2Å(更准确些:C—H、N—H 和 O—H 约 1Å,C═O、C ═ O、C ═ N 和 C═C 为 1.2~1.3Å,C—C 约为 1.5Å,S—S 约 1.8Å)。

键角的典型值约为 120° 和 109°。120° 角见于 sp^2 杂化的 —C′ 、—N 原子,在这些位置三个共价键伸出指向三角形的三个顶点,而 109° 角见于 C^α 型的 sp^3 杂化原子,在该处四个键伸出指向四面体的顶点,在各有两个键的 O 或 S 原子上也是 109°。

■ 现在来看一下涨落(即键结和键角的热振动)的特征值。它们可能对蛋白质链的柔性有贡献。

振动频率显示在蛋白质的红外谱(infrared,IR)中。特征振动频率如下:氢原子的振动,如在 C—H 键中(相应的 IR 光的波长为 $\lambda = c/\nu \sim 5\mu m$)$\nu \sim 7\times10^{13}\ s^{-1}$;"重"基团,如 CH_3—CH_3键的振动 $\nu \sim 2\times10^{13} s^{-1}$(此时 $\lambda = c/\nu$,15μm;其中 c 为光速,300 000km · s^{-1})。

在室温下这些振动是否可以激发?

要回答这个问题,我们必须把每一个自由度的热能("热量子"kT)与振动能量进行比较。来估计一下在常温下的 kT 值。其中 T 为绝对温度,单位为开尔文(在 27℃,即大

致的室温时，$T=300\text{K}$。K 表示开尔文)。而 k(有时也写作 k_B)为波尔兹曼常量(等于每摩尔 2cal/(mol · K)，或每个粒子 0.33×10^{-23} cal/K，因为每摩尔含有 6×10^{23} 个粒子)。因此，在室温下"热量子"$kT = 600\text{cal}/(6\times10^{23}$ 个粒子)，即每个粒子 10^{-21} cal。

与这一热量子相应的频率 ν_T 可以由关系式 $kT=h\nu_T$ 得到[其中普朗克常量$h\equiv2\pi\hbar= 6.6\times10^{-34}$ J · s = 1.6×10^{-34} cal · s。提醒一下，1cal(卡) ≈4.2J(焦)]。于是在 $T=300°\text{K}$，即 27℃时，频率 ν_T等于 $7\times10^{12}\ \text{s}^{-1}$。

"热量子"不能激发更刚性的，即频率比它自身更高的振动。

因此，在室温下共价键是"硬的"，不会振动：它们的振动频率 ν 为 2×10^{13} ~ $7\times10^{13}\ \text{s}^{-1}$，即它比 $\nu_T= 7\times10^{12}\ \text{s}^{-1}$要高一个数量级。

然而，共价键可以因红外光照射而振动，这是蛋白质红外光谱学(IR)的基础。实际上，正是 IR 提供了关于原子、共价键和共价键角的信息。当然，对它们的实验研究是从小分子开始的，然后所获得的信息应用于蛋白质研究中。

与键长振动不同的是，键角的热振动要"软"一个数量级(它们的频率在 10^{12} ~ 10^{13}范围内)，在室温下可以被激发。但是它们的特征振幅并不大：大约 5°。

因此，键长的振动对蛋白质链的柔性完全没有贡献，而键角振动的贡献也很微薄。本质上，蛋白质链的柔性，即它能折叠成 α 螺旋或球体的能力，是由**围绕**价键的转动(虽然不是完全自由的旋转，见下文)所赋予的。因此整个链的结构常常简单地以围绕价键进行转动的角度(这被称为"构象")来描述。因此，术语"结构"和"构象"常常被简单地作为同义词而使用。

跟前面一样，原子基团围绕价键旋转的主要信息也是由红外光谱学提供的。我只给出结果。

■ 由价键相连的原子基团的相对位置以二面角(因为该角处于两个平面之间：在图 2-3 中，a、O′ 和 O″处于一个平面中，O′、O″和 c 处于另一个平面中)来描述。

图 2-2 和 2-3 图示说明了二面角的读数方式。读数方式与中学里的三角学一样：如果把离我们较近的价键看做是"转动箭头"，较远的价键为"坐标轴"，中心价键为"转动轴"，那么"箭头"逆时针转动时转动角增加，顺时针则转动角减小。

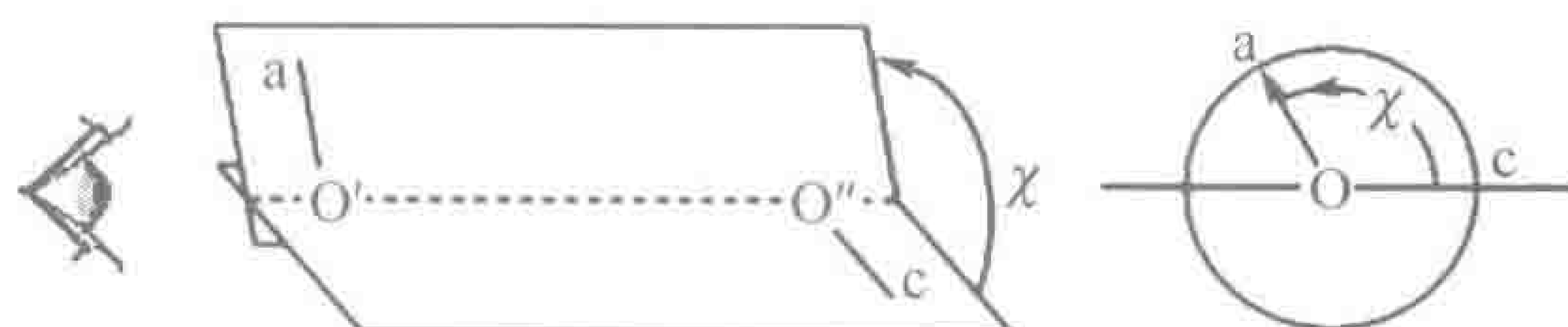


图 2-3　二面角(转动角)的读数：横向视图(左)和垂直旋转轴的截面视图(右)。离读者较远的化合键 O″—c："坐标圆的轴"；中心连接 O′—O″："转动轴"；离读者较近的连接 O′—a ："坐标圆中的箭头"

■ 图 2-4 显示了在绕由两个 sp^3 杂化原子结合而成的价键(以 $H_3C—CH_3$ 和 $CH_2C—CH_2C$ 键为例)转动时的能量特征信息。这种价键对脂肪族侧链基团是很典型的。侧链基团的转动角被称为 χ ("chi")角(图 2-2)。这种三重态(对应于绕

sp^3-sp^3键的转动对称性)势能(即在旋转 360° 时势能具有三个最大值和三个最小值)的最大值对应于导致电子云靠近(并排斥)的三个"重叠"构象(0°,120° 和 -120°)。由于形成这些电子云的电子已经处于共价键中了,所以电子云互相排斥。因此而出现的绕 $H_3C—CH_3$ 价键转动势垒的值约为 3kcal/mol,而在最小值附近的热波动(即能量增大 kT 时的偏差)的典型范围是 15~20°。

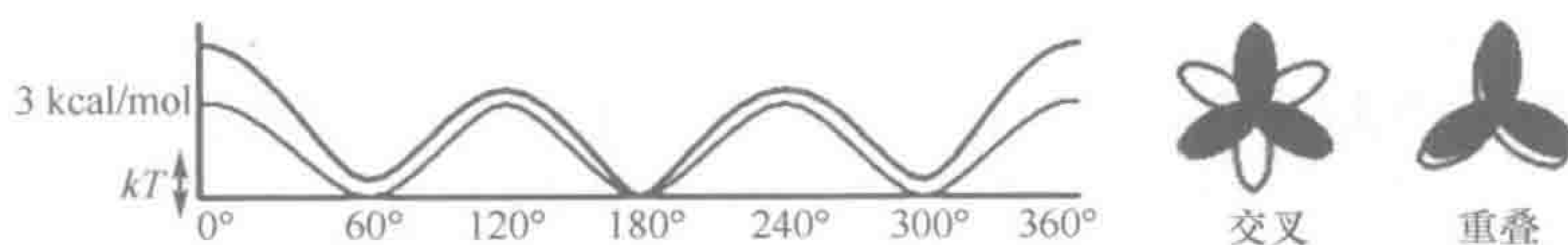


图 2-4　围绕相同的在两个 sp^3 杂化原子之间的价键转动时的典型转动势能曲线:围绕 $H_3C—CH_3$ 键(下方曲线)和围绕($C—CH_2$)—($CH_2—C$)键(上方曲线)。主要的能量效应来自于电子云的排斥:在"重叠"构象时能量最大(0°,120°,-120°),在"交叉"时最小(60°,180°,300°)。小 H 原子的排斥并不重要。但是被大电子云包围的重 C 原子在 0°(C—C—C—C 链上)附近的排斥产生了额外的能量效应,使得绕($C—CH_2$)—($CH_2—C$)键与绕 $H_3C—CH_3$ 键的转动有区别。"↕"箭头标示了 kT 值以便对比

当占据 sp^3 杂化原子的几个共价键的不是氢原子,而是更重的原子时,它们的排斥力会增大与这些原子过于紧密接近处所对应的势垒部分。这以($C—CH_2$)—($CH_2—C$)中绕中心价键转动为例显示在图 2-4 中。

图 2-5 中显示了在围绕由两个 sp^2 杂化原子(C′和 N)结合而成的肽键转动时的能量特征信息。围绕这种键的转动角被称为 ω 角(图 2-2)。势能是二重态的(即在旋转 360° 时具有两个最大值和两个最小值),对应于绕 sp^2-sp^2 键的转动对称性。由于 p 电子参与了肽键(如本节开头所讲到的),势垒很高。势能最小值出现在 0° 和 180°(该处拉住 C′ 和 N 原子的 p 轨道靠得最近),而最大值在 90°和 270°(该处 p 轨道离得最远,相应地,相互结合最差)。由于势垒很高,围绕这种价键转动的热波动的典型范围并不大:约 10°。

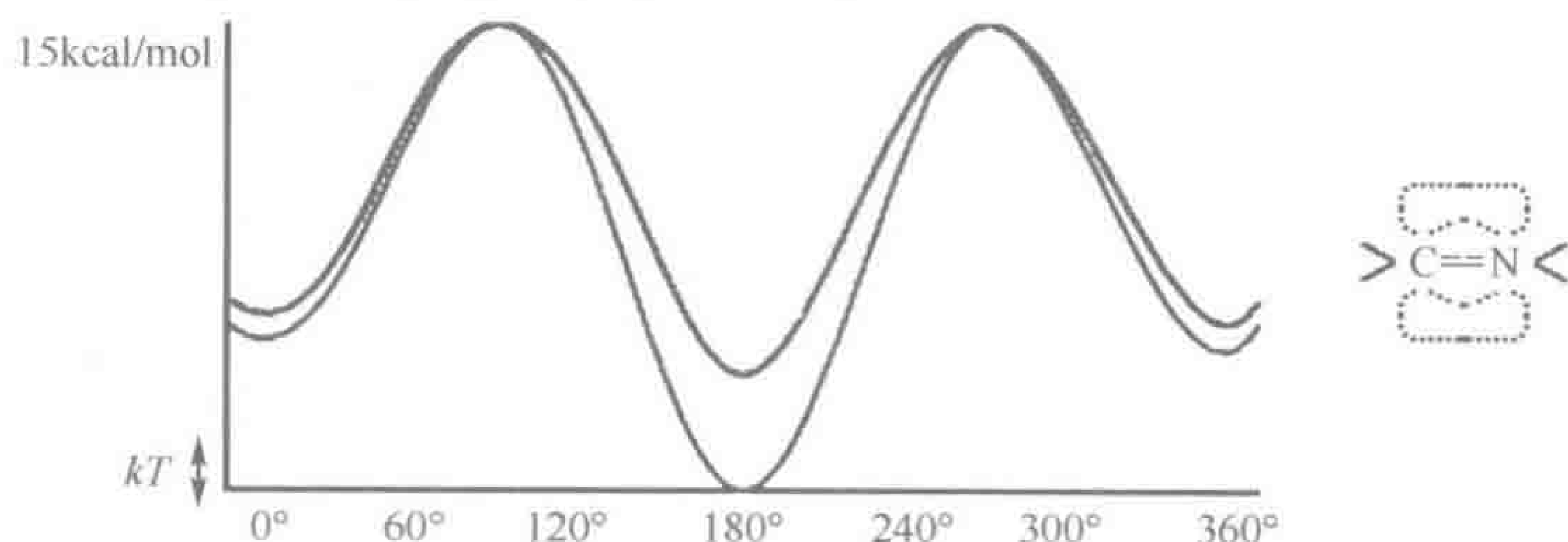
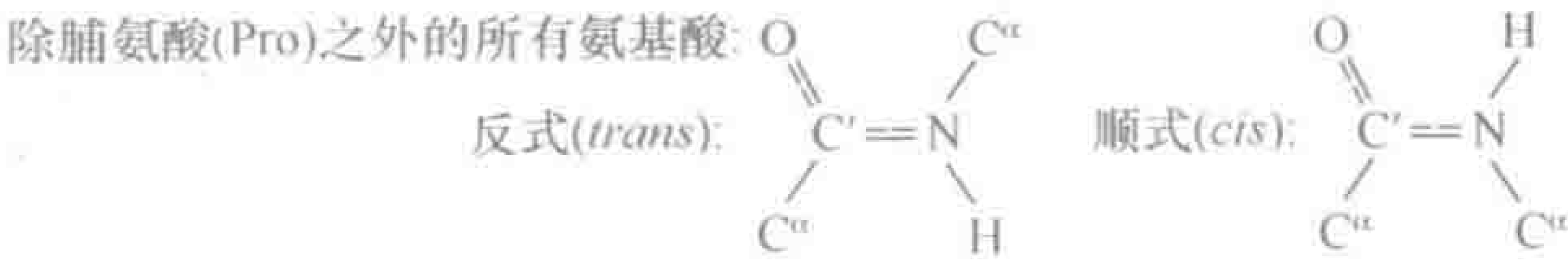
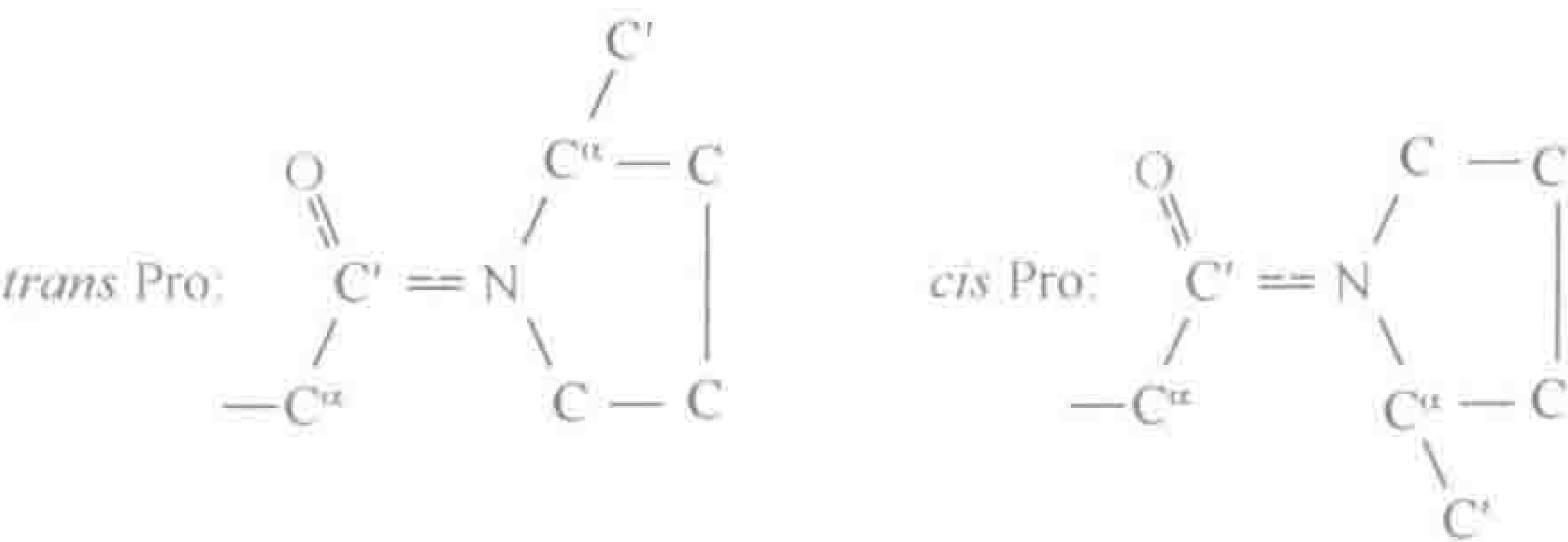


图 2-5　肽键中围绕连接两个 sp^2 杂化原子(C′和 N)的短键的典型转动势能曲线。右边显示了电子 p 轨道在角度等于 0°或 180°时的键。对于所有肽键(在脯氨酸之前的除外),由于在 0°处存在重 C^α 原子的排斥,在 0°的能量高于 180°的能量(下方曲线)。位于脯氨酸之前的肽键(上方曲线),由于在 N 原子上的不是一个,而是两个 C 原子,所以这种能量差别较小。参见下文的文字和图片,其中说明了脯氨酸的结构形状

值得注意的是，重 C^α 原子的排斥作用使顺式（*cis*）构象（$\omega=0°$）在能量上颇为不利；因此在蛋白质中的几乎所有肽基团都处于反式（*trans*）构象（$\omega=180°$）。



例外的是在脯氨酸（Pro）（它是亚氨基酸，而非氨基酸）之前的肽键：脯氨酸的 N 原子有两个（除开 C′原子之外）几乎对称的重原子团（—$C^\alpha HC_2$ 和—CH_2C，图 2-1），因此对这个键反式构象并不比顺式构象好多少。



在球蛋白和解折叠的（即变动的，不具有固定结构的）多肽中，观察到脯氨酸有 90% 为反式，10% 为顺式。我希望你们注意到这个规律：对自身越有利的，在球蛋白中就越经常出现。我们将会不止一些碰到这个规律。

最后，来看看围绕由 sp^3 杂化和 sp^2 杂化原子连接的价键的转动势能。绕这种键的转动角被称为 ϕ 角（绕 N—C^α键）和 ψ 角（绕 C^α—C′ 键）（图 2-2）。绕它们任何一个转动都出现了六重态（转动 360° 有 6 个最小值和 6 个最大值）势能，势垒相当低（图 2-6）：其高度约等于 1kcal/mol，即热振动能量的数量级（我们记得，在室温下是 0. 6kcal/mol）。正是绕多肽主链中的这种价键的 ϕ 角（在 N 与 C^α之间）和 ψ 角（C^α与 C′ 之间）的几乎自由的转动提供了蛋白质链的主要柔性。

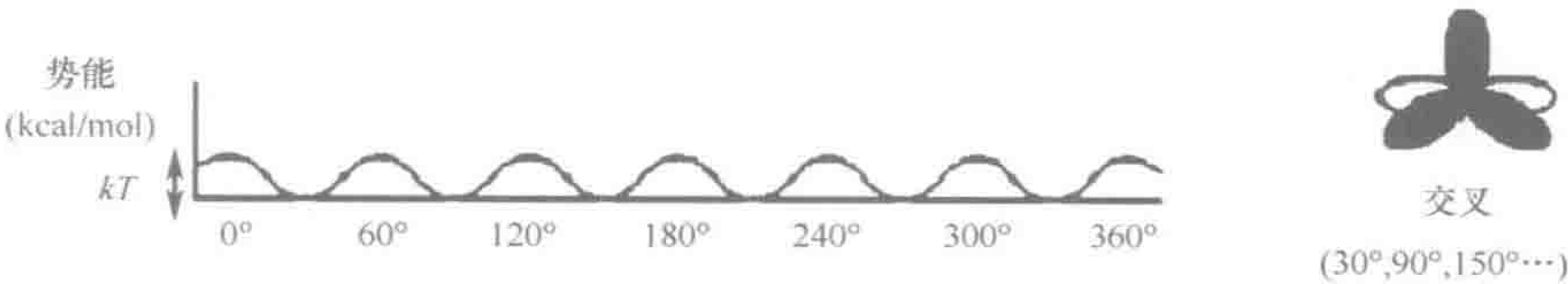


图 2-6 绕 sp^3 杂化和 sp^2 杂化原子间的价键的典型转动势能曲线（以绕 $H_3C—C_5H_6$ 键转动为例）

第3讲

范德华相互作用:远距时吸引,近距时排斥。氨基酸残基的允许构象(甘氨酸、丙氨酸、缬氨酸和脯氨酸的拉氏图)。

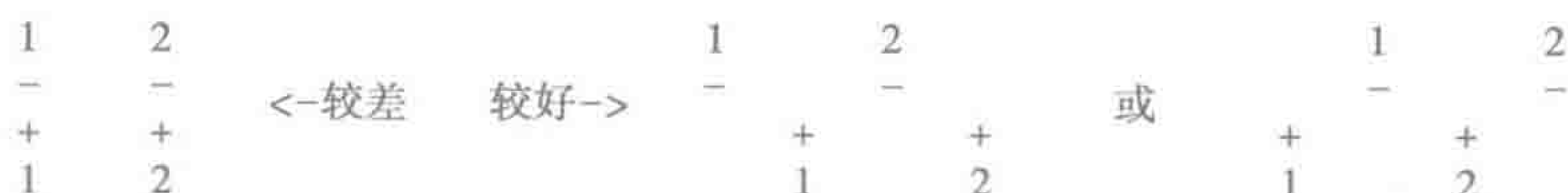
现在来讨论彼此之间没有共价键相连接的原子之间的基本相互作用。

不过让我们首先来回顾一下,在什么情况下相互趋近的原子之间并不形成共价键。我提醒一下,形成共价键的基本原因是电子的离域化,即它们从一个原子来回“转移”到另一个原子,或者,正如通常说的,电子驻留在包围两个或更多原子的共存的“共有”轨道。不过,根据**泡利不相容原理**,在同一位置,同一轨道中只能够容纳不超过两个电子,而且同时它们的“自旋”(即电子的转动动量)必须方向相反(或者说这两个电子“成对”)。两个电子在一个共存轨道上的成对(一个电子来自一个原子,另一个电子来自于另一个原子)会导致形成牢固的共价键。

如果两个趋近的原子的每一个轨道都有两个成对电子,则这些原子不能形成共价键。否则在每个共有轨道上有过多的电子:四个。而根据**泡利不相容原理**,每个轨道最多容纳两个!这意味着,已经饱和、填满电子对的轨道不能进一步公有化。当两个带有这样的“饱和”轨道的原子靠近到使它们的电子云开始重叠时,它们就会排斥。在通常的、并不超高的温度下,这些原子不能互相穿透。

如果靠近的分子上所有的化合价都已饱和,则会出现同样的情况:如果它们的原子靠近到2~3Å时,它们就会相互排斥。

但是在相距较远时(电子云不发生重叠),则只有当它们不带电时(对带电情形我们后面将会讲到),所有原子和分子都相互吸引。这种吸引纯粹是量子性的。这与两个原子中的电子的相干振荡相关。原因在于,电子的协同(“在同一个方向”)位移会导致原子的互相吸引(反之,电子无偏移的原子既不相互吸引也不相互排斥)。从下面示意图中可以清楚地看出(图中有两个原子,“1”和“2”,每一个都含有电子“-”和核“+”):



观察这个示意图时,应该记得,静电能是按 $1/r$ 关系随距离 r 的减小而增大的。电子移动的结果并不改变电子与电子、核与核的相互作用,但是(请看最右边的图片)电子1对核2的吸引力增强了,而且它比电子2对核1的吸引力减弱要大得更多。现在请记住,原子中的电子总是在振动(“绕核转动”,而且由于海森堡量子不确定性而不会掉到原子核上去)。上述效应能够激起相互靠近的原子出现电子的相干振荡,从而引起这些原子间的吸引。

在原子远离时，它们之间的相互作用减弱。结果是相互作用能量随核间距离 r 而减小。可以证明，它正比于$(1/r)^6$。

原子的总相互作用能量，也称为范德华相互作用能量，显示于图 3-1 中，近似地可用勒让德-琼斯(Lennard-Jones)公式描述：

$$U_{LJ}(r) = E_0\left[\left(\frac{r_0}{r}\right)^{12} - 2\left(\frac{r_0}{r}\right)^6\right] \tag{3.1}$$

式中，r_0 为 U_{LJ} 处于最小值时的距离(取 U_{LJ} 对 r 取导数即可很容易地证明)。而$-E_0$ 是此最小值的深度。公式(3.1)的最后一项以（常数/距离）的 6 次方而随距离减小，对应于吸引力(它前面的负号表示在靠近时相应的能量降低)；而 12 次方的那一项对应于排斥力(此项为正数，即在靠近时相应的能量升高)。

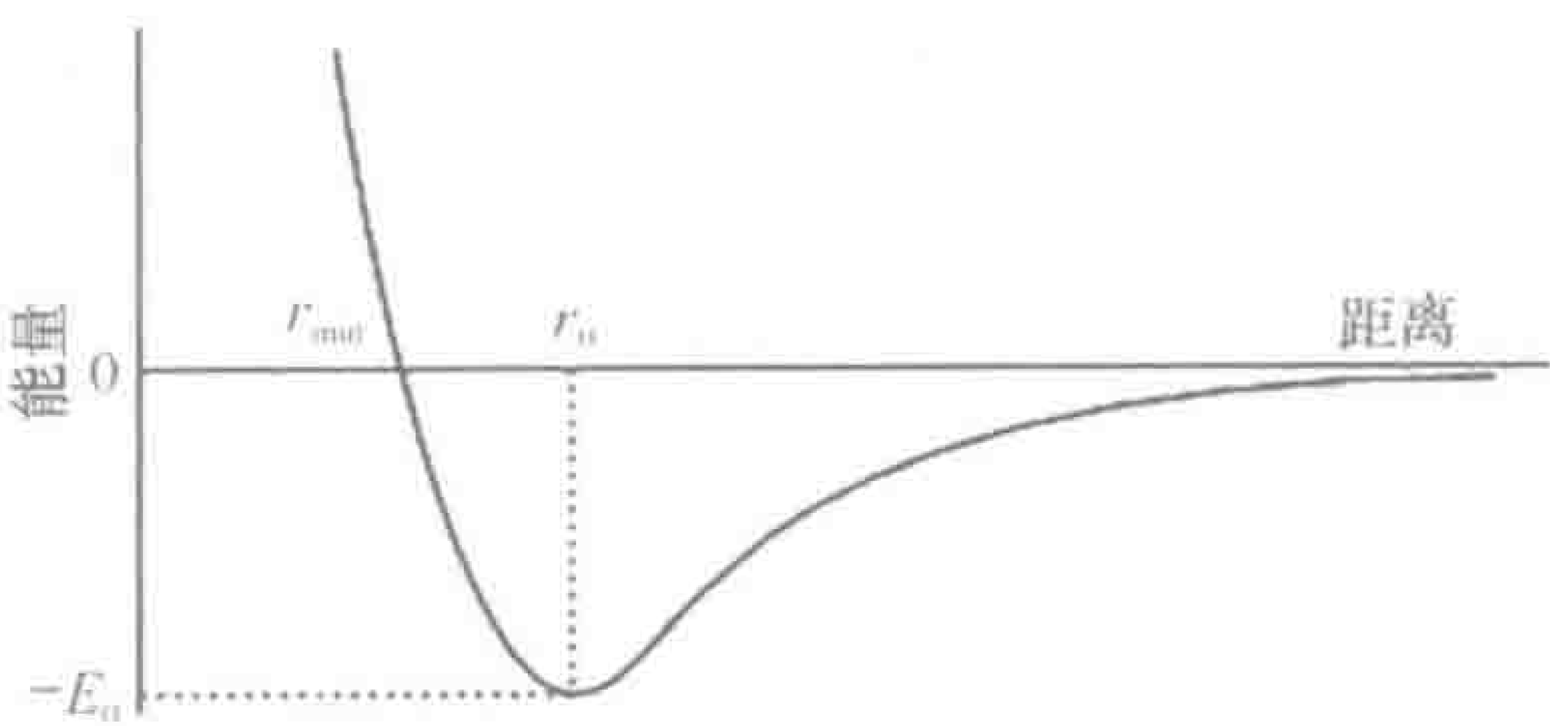


图 3-1　范德华相互作用势能的典型形状

公式(3.1)准确描述的仅仅只是在远距离(当 $r>>r_0$ 时)的吸引力。在近距离时的排斥力仅仅只是定性地描述为“当 r 趋近于零时，排斥力极强，远远超过任何吸引力”。公式(3.1)是近似性的，可以从它暗含着原子的球状性(因为它所描述的相互作用不依赖于方向)看得出来。而原子的电子云，通常来说，不是球状的(请回想一下向外伸展的 p 电子)。一般来说，只有量子力学才能完全正确地描述原子的相互作用，但是它能精确计算的只有极其简单的系统，如 He 原子、H_2^+ 离子或 H_2 分子。其他系统只能取如公式(3.1)这样“半经验的”的近似公式。其中公式的形式取自定性的物理学考虑，而参数(我们的 E_0 和 r_0)来自实验。几个基本数据列于表 3-1 中。

表 3-1　范德华作用势能的典型参数

相互作用	E_0/(kcal/mol)	r_0/Å	r_{min}/Å	原子的范德华半径/Å
H ⋯H	0.12	2.4	2.0	H:1.0
H ⋯C	0.11	2.9	2.4	
C ⋯C	0.12	3.4	3.0	C :1.5
O ⋯O	0.23	3.0	2.7	O :1.35
N ⋯N	0.20	3.1	2.7	N :1.35
CH_2⋯CH_2	≈0.5	≈4.0	≈3.0	CH_2:≈1.5

注：原子间相互作用的 E_0 和 r_0值引自 Scott R. A. ,Scheraga H. A. 1965. J Chem Phys. 45:2091;r_{min}值引自 Ramachandran G. N. ,Sasisekharan V. 1968, Adv Prot Chem. 23:283. CH_2⋯CH_2 相互作用的参数由这些数值估计而得。由于 CH_2⋯CH_2 的相互作用依赖于这些基团的相对取向，因此表中的数值是近似的；然而在计算蛋白质中的相互作用时经常需要使用它们，因为 X 射线“看不见”蛋白质中的氢原子。

请注意，表中不仅给出了最佳距离 r_0，还给出了 r_{min}，即在晶体中观测到的相应

原子的最小距离。r_{min}大致相当于图 3-1 中的在原子间距离极短时能量通过 0 的点。r_{min}值在估算某种肽链构象是否可能时很有用。

课堂里的声音：已经有大量对原子间相互作用势能进行推导的工作。是否意味着，不能相信这些势能的准确性？

讲课人：对于势能的**形状**(图 3-1)没有什么意见分歧。对于r_{min}的估计值也没有特别的分歧，它们是直接从晶体结构中获得的。对参数r_0，特别是E_0的值存在有分歧……距离r_0和能量E_0基本上也是从晶体中[它们的结构(距离)和升华热(能量)]获得的。但晶体通常不是由原子组成的，而是由分子组成的，如CH_4，C_2H_6，…，因此在计算能量时总是出现这种问题：C…C 相互作用的贡献如何，H…H 如何，C…H 如何？不同的作者用不同的方法来解决，因此它们的势能有时会有明显差别。然而，在由原子反向过渡到分子时，这些差别通常被平滑处理掉了。但是，如果你需要计算分子结构，必须记住，绝对不能从一个作者那里取一组参数(例如，C…C 相互作用的能量)，而从另一个作者那里取另一些(例如，H…H相互作用的能量)：这里必须遵守“全或无”原则，否则会出错。

列于表中的数值有助于理解为什么 C′══N 键容许有反式(*trans*)构象却不容许顺式(*cis*)构象(如前所述，对于除脯氨酸之外的所有氨基酸残基)：在 C′══N 键的反式构象中 C^α 原子间的距离为 3. 8Å，而在顺式构象(此时这些原子靠得最近)中总是 2. 8Å，这小于所允许的最小距离r_{min}(在 C…C 中等于 3. 0Å)。



■ 由于反式 C′══N 键的刚性，而且由于里面沿链相邻的氨基酸的 C^α 原子彼此相距比较远，沿链相邻的残基可以几乎彼此独立地改变自己的构象。但是残基内部绕 ϕ 和 ψ 角的转动是相互关联的。残基的“允许”构象和“禁止”构象的图像可以按照坐标 (ϕ，ψ) 绘制，这种图被称为**拉氏图**①(Ramachandran 图，或更准确些：Sasisekharan Ramakrishnan-Ramachandran 图)。

把绘制这些图之前，让我们看一看，在单独沿 ϕ 角(绕 N—C^α键)和 ψ 角(绕 C^α—C′ 键)转动时，哪些构象是允许的，哪些是禁止的。

我们已经知道，绕这些键(连接 sp^3杂化的 C^α原子和 sp^2杂化的 N 或 C′ 原子)的转动实际上是自由的。然而在顺式构象中($\phi=0°$ 或 $\psi=0°$ 时)绕这些键转动的原子(在绕 N—C^α键沿 ϕ 角转动的 C'_{i-1}和 C'_i，绕 C^α—C′ 键沿 ψ 角转动的 N_i和

---

①中文完整译名为“拉马钱德朗图”，简称“拉氏图”。参见文献：Ramachandran G. N. , Ramakrishnan C. , Sasisekharan V. 1963. Stereochemistry of polypeptide chain configurations. J Mol Biol. 7：95-99.

N_{i+1}。如图 3-2 所示；i-1，i，i+1 为链上连续残基的编号）彼此非常接近，此时由于这些原子的排斥，构象可能被禁止，或者说，在立体结构上是被禁止的。

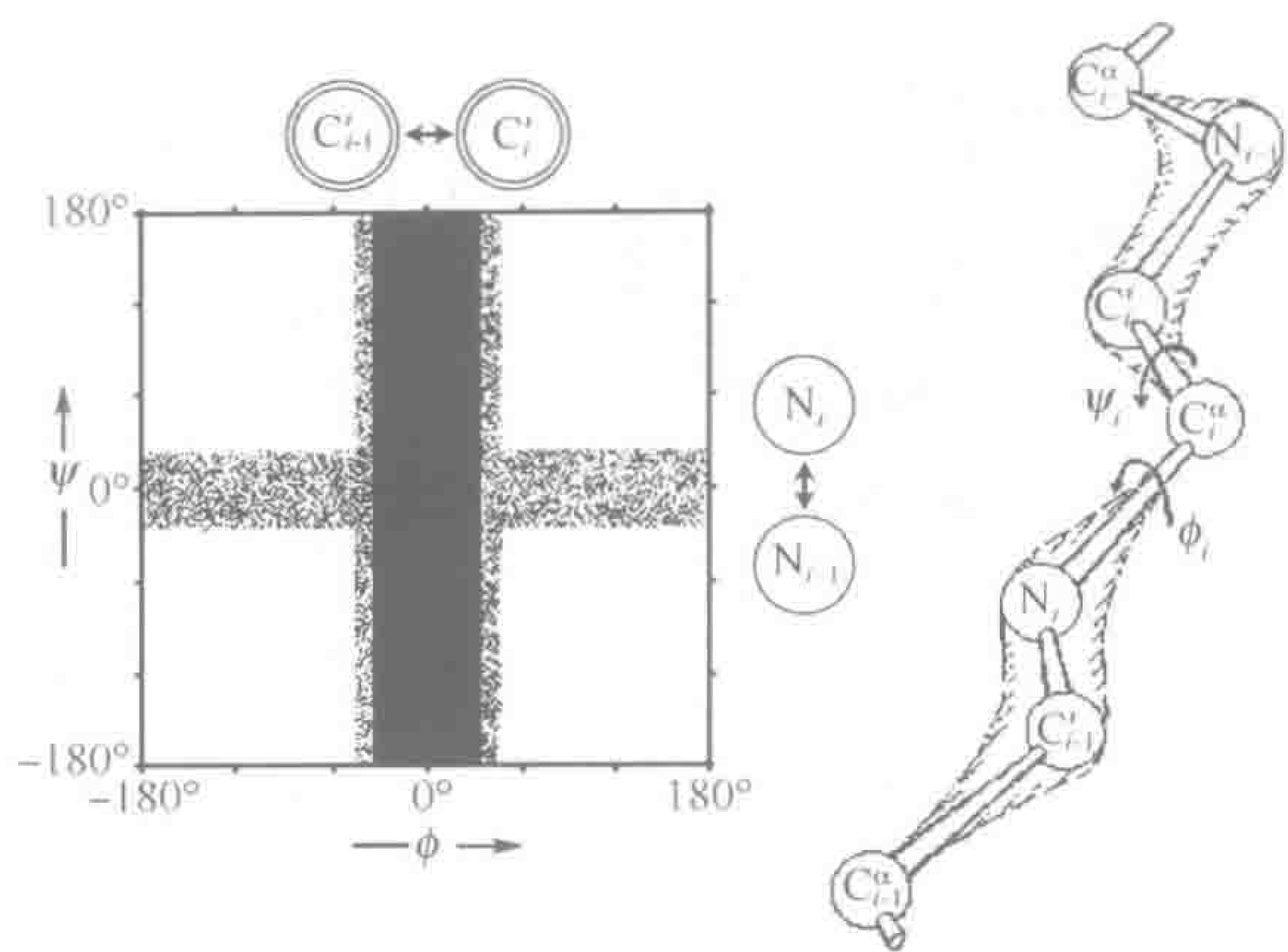
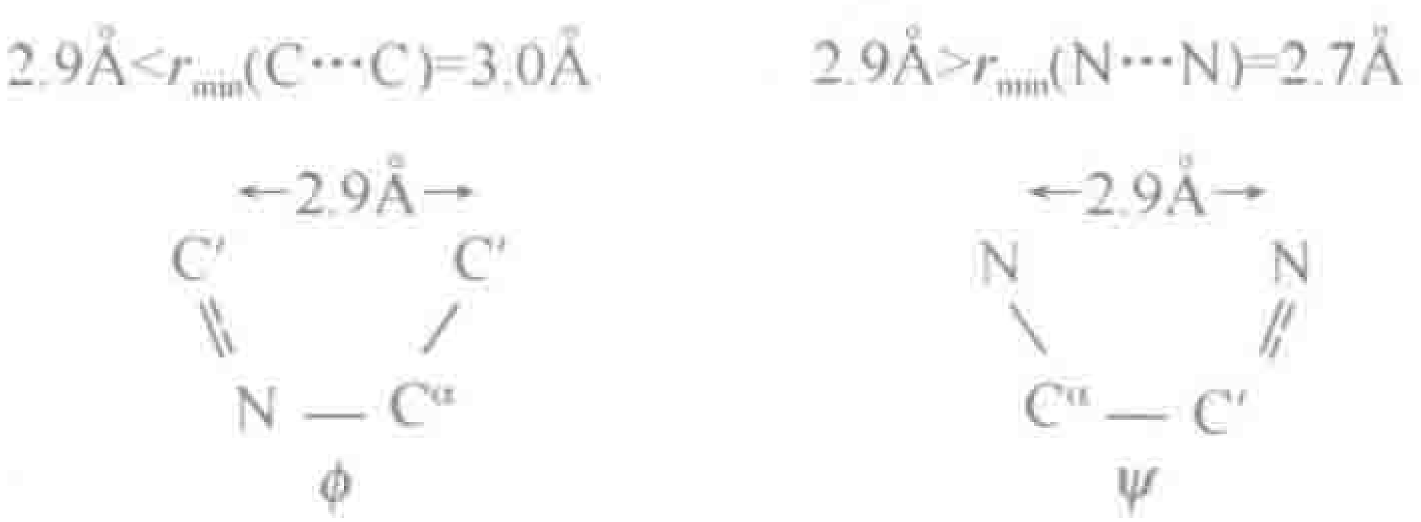


图 3-2　在片段 $C^{\alpha}C'N—C^{\alpha}—C'NC^{\alpha}$ 中，如果这些原子上没有连接更多其他原子（右图），且残基 i-1 与 i+1 上的原子间不发生相互作用，则 ϕ 和 ψ 角转动时的禁止构象(■)、应变构象(▒)和完全允许构象(□)的形状

为什么沿 ϕ 角的转动比沿 ψ 角的转动更困难呢？

如下示意图表明，C'_{i-1}与 C'_i（沿 ϕ 角转动时）、N_i 与 N_{i+1}（沿 ψ 角转动时）之间的原子最小距离是一样的，为 2.9Å，比 C…C 相互作用的 r_{min} = 3.0Å 略微小一点（因此，沿 ϕ 角转动时，顺式构象被禁止），比 N…N 相互作用的 r_{min} = 2.7Å 略微大一点（虽然比最佳的 r_0 = 3.1Å 还是小一些）（所以顺式构象未被禁止，但是是受应力的）。



而且，计算显示，共价键 $N—C^{\alpha}—C'$ 所具有的甚至很小的柔性都能强烈地降低这种应力的能量。

如果只限于考虑这些 C′ 和 N 原子，则禁止构象、应变构象和允许构象图就如图 3-2 所示，在这种情况下沿 ϕ 角和 ψ 角的转动是彼此独立的。

然而在 C′ 原子上有 O 和 C^{α}原子，而 N 原子也有 H 和 C′ 原子（而且在水中，正如我们在下一讲中将会看到的，在 H 原子上通过氢键牢固地与 H_2O 分子相连）。因此结果是，甘氨酸（Gly）（最小的氨基酸残基，侧链基团为 H）的拉氏图：禁止和允许构象图，如图 3-3 所示。

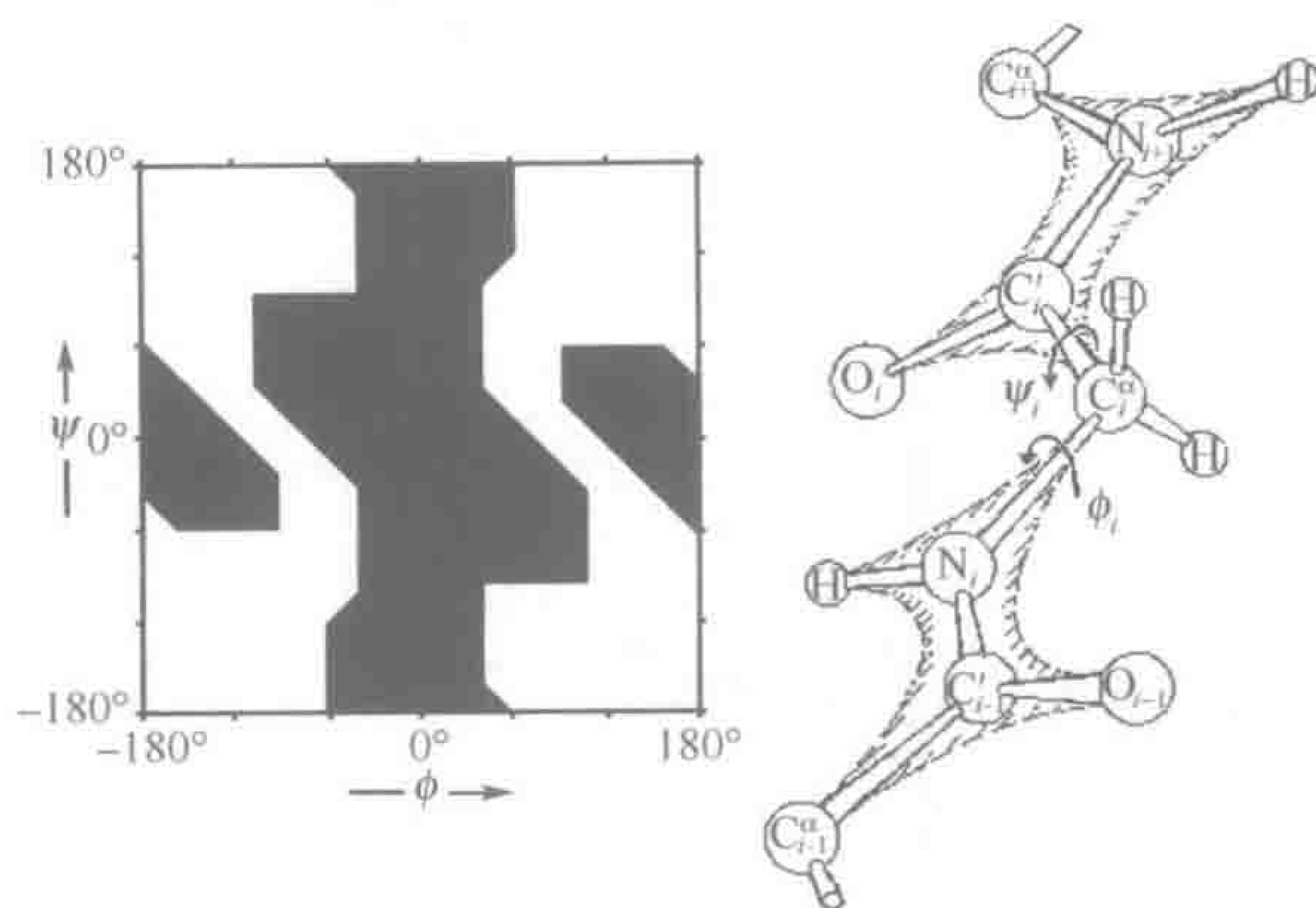


图 3-3　在蛋白质链中绕 ϕ 和 ψ 角转动时甘氨酸(Gly)的禁止(■)和允许(□)构象图

甘氨酸没有重侧链原子团,而所有其他氨基酸残基都有,这些原子团(更准确地说,其靠近主链的 C^{β}原子)与 C'_{i-1}原子的碰撞裁剪出沿 ϕ 角的禁止区域,与 N_{i+1}原子的碰撞裁剪出沿 ψ 角的禁止区域(图 3-4)。

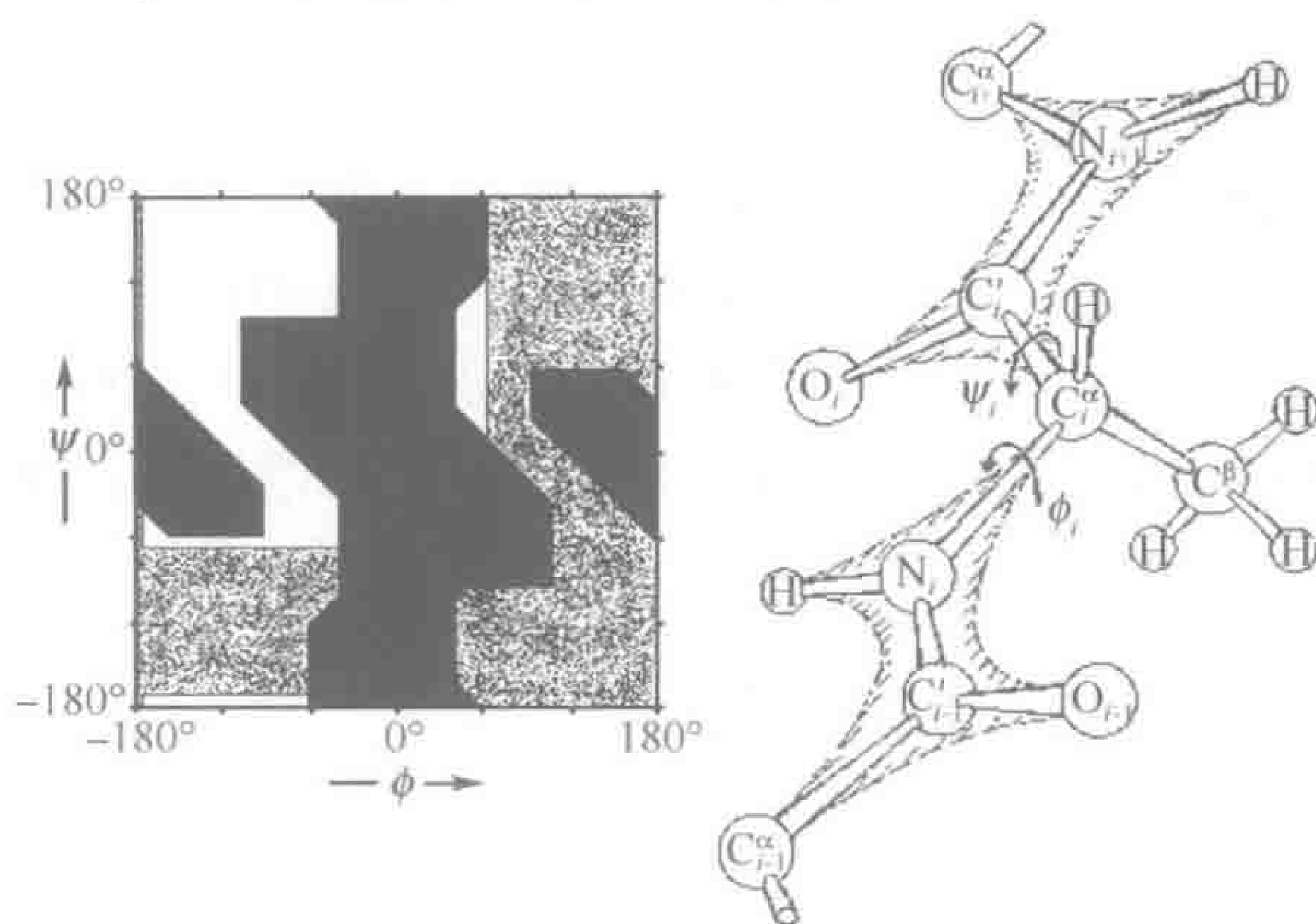


图 3-4　在蛋白质链中绕 ϕ 和 ψ 角转动时丙氨酸(Ala)的允许(□)构象图;(▒)为仅对甘氨酸允许的区域;(■)为对所有残基的主链相互作用都禁止的区域

■　图 3-4 中所示的是丙氨酸(Ala),其侧链原子团较小:$C^{\beta}H_3$基团。其他氨基酸残基的原子团都更大:在 C^{β}原子上都连接有一两个重 γ 原子。然而这些"新的" γ(以及更远的 δ,ε 等)原子离主链较远,因此对拉氏图影响很小。更确切地说,在图 3-5 中的白色小区域内这些 γ 原子完全不会与主链相碰撞;而在丙氨酸其他允许构象区域这样的碰撞对原子团的某些构象是有可能的,因此这里并不是对原子团的所有构象都允许;这些区域我用较淡的阴影绘于图 3-5 中。

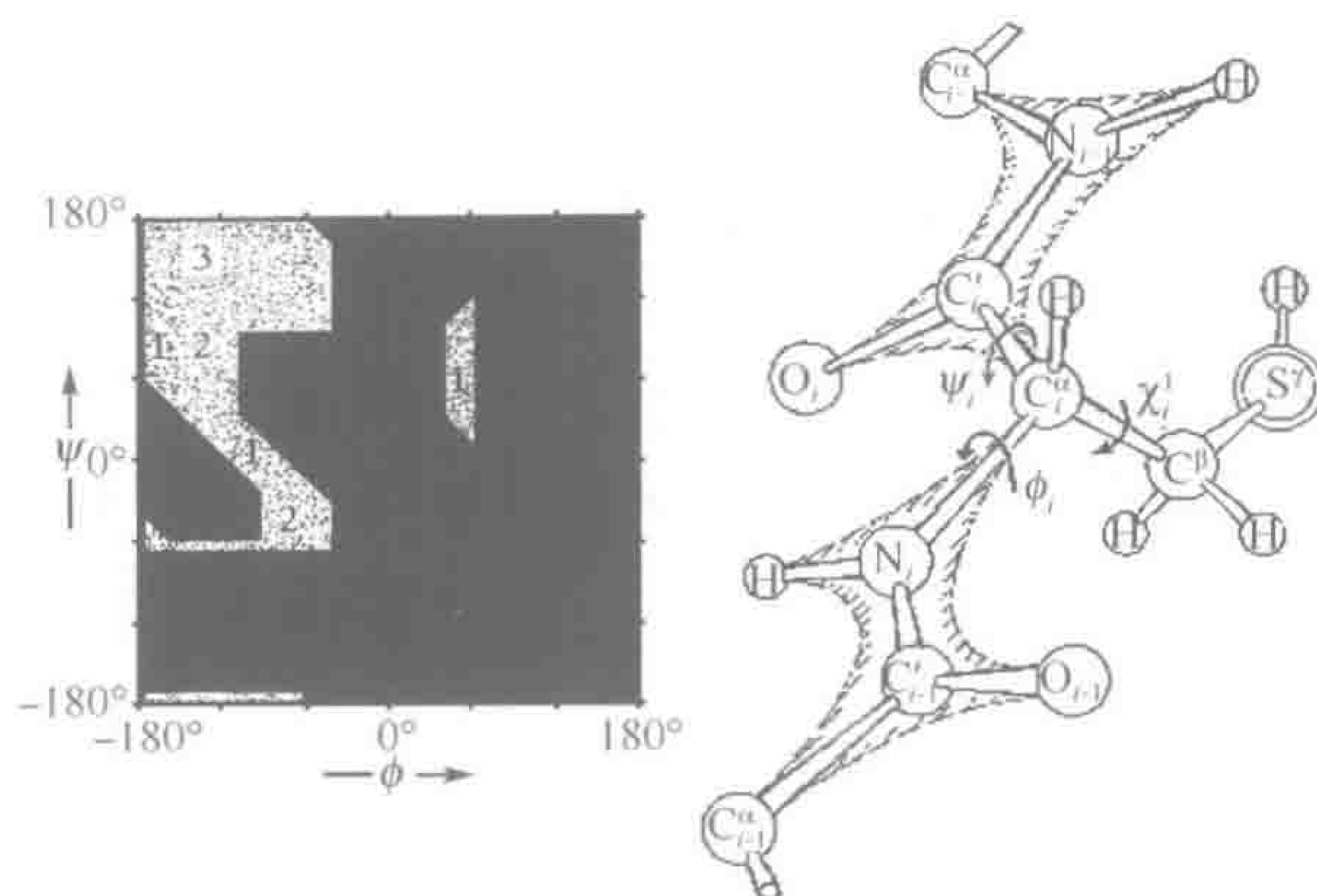


图 3-5　在蛋白质链中沿 ϕ 和 ψ 角转动时较大残基的禁止(■)和允许(□,▒,▓)构象图。在□区域是对 C^{γ} 原子的所有 3 种 χ^{1} 旋转异构体都允许。▒区域允许 2 种,▓区域只允许 1 种

γ 原子与主链的碰撞对于缬氨酸(Val)、异亮氨酸(Ile)和苏氨酸(Thr)特别重要,因为它们带有两个大 γ 原子。

■　最后来看亚氨基酸脯氨酸(Pro)的拉氏图。它的 ϕ 角几乎是固定在−70°,这是因为它有一个由侧链基团形成的环连接到主链上的 N 原子;而绕 ψ 角的转动与丙氨酸一模一样。因此它的允许构象看上去就是拉氏图中的白色条带,如图 3-6 所示。

需要注意的是,脯氨酸的环会压缩位于环上前一个残基的允许构象(图 3-7)。

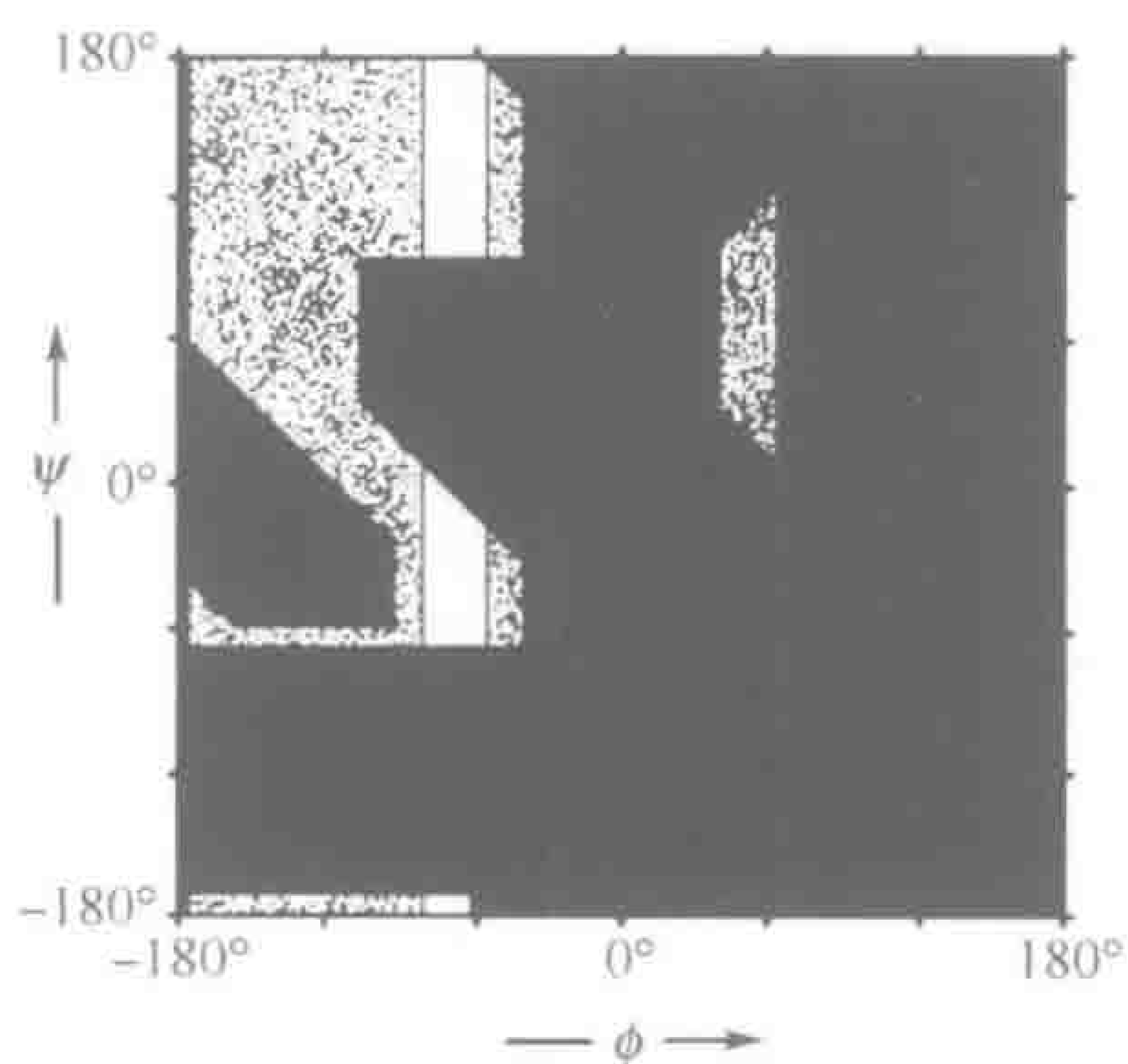


图 3-6　以丙氨酸(Alα)允许构象(▓)为背景的脯氨酸(Pro)的允许构象图(□)。■为对两者都禁止的构象区域

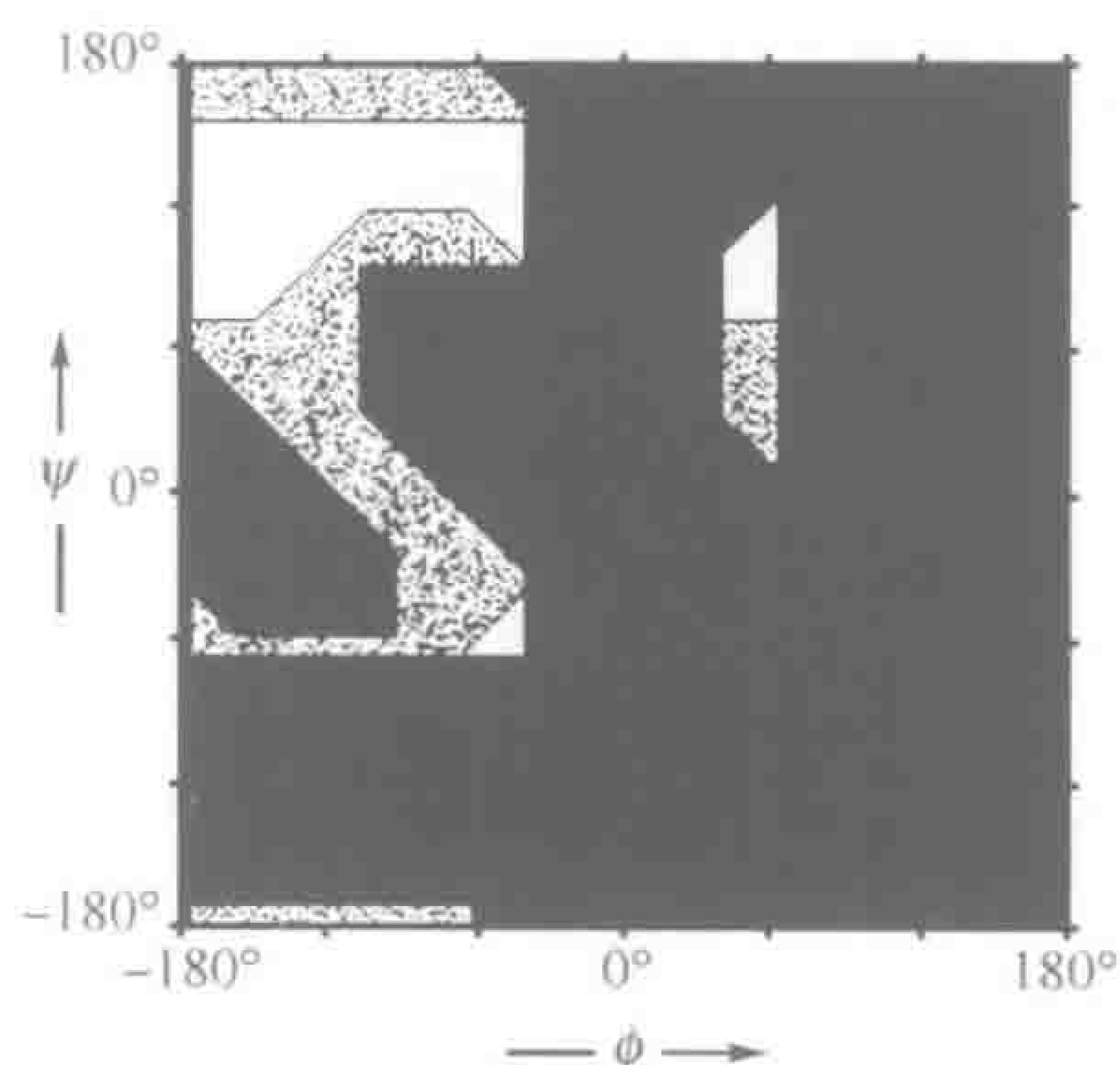


图 3-7　在蛋白质链上位于脯氨酸之前的丙氨酸的允许构象区域(□)。如果没有脯氨酸,(▓)区也是允许构象区域

■　最后,让我们来看这些理论计算结果与在蛋白质中观察到(X 射线和核磁共振

谱)的氨基酸残基构象符合得怎么样。在图 3-8 里,一边以理论计算对甘氨酸允许区域的轮廓线为背景,另一边以理论计算对丙氨酸和其他残基允许区域的轮廓线为背景,实验中观察到的构象以点绘制在图中。

看得出来,符合得非常好。

绝大多数实验点落于“立体结构上允许”的区域。但是有一些点落在“立体结构上禁止”的区域。对此不必惊讶,因为我们称为“立体结构上禁止”的是那些仅仅能量较高的区域。当然,并不是无限高,以致完全禁止。例如,比能量最小的构象高出大约几个 kcal/mol。即,如果要把氨基酸残基赶入那些区域,则蛋白质必须耗费能量,但是它完全可以做到。可以做到,不过很少这样做。

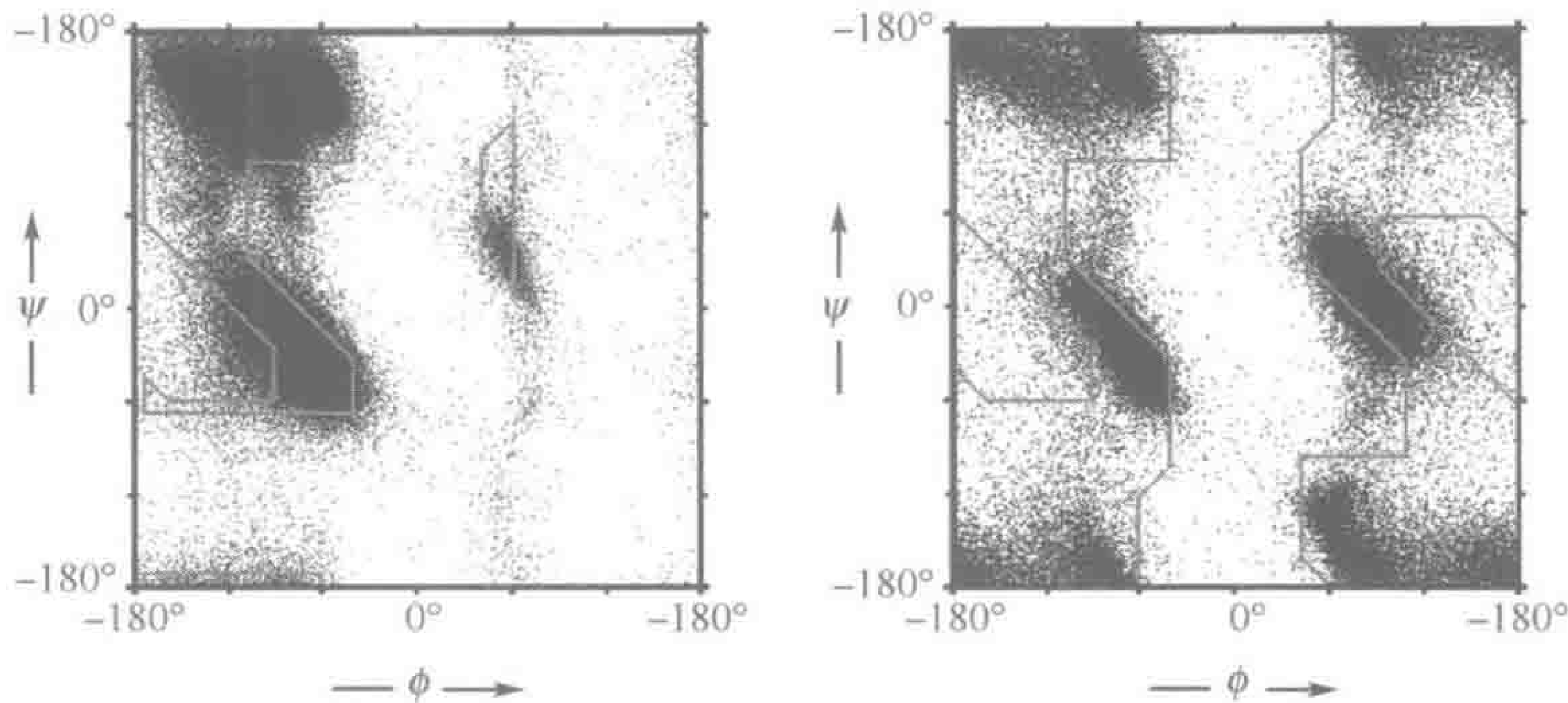


图 3-8　在蛋白质中观察到的甘氨酸(右)和所有其他残基(左)的氨基酸残基构象(点)。轮廓线描出了立体结构上的允许区域

在本课程的学习过程中,我们将会理解这样一个普适原则:处于应变状态的、高能量的成分在蛋白质中很少碰到,但是总能碰到。不需惊讶的是:一旦蛋白质稳定下来,它应该(主要地,但并非绝对完全地)由稳定的元件所构成。

第4讲

水环境的影响。氢键。氢键的电本性。氢键的能量。氢键在晶体中的几何形状。氢键在水中的松弛。熵和自由能的概念。肽链中的氢键取代了该链与水的氢键结合;结果使得肽链中的氢键获得了(在水环境中)熵性质。

直到现在为止我们还没有考虑蛋白质的水环境。现在是填补这个空白的时候了。

水,是非常特别的溶剂。首先,对其低相对分子质量而言,它的沸点和凝点是不同寻常的高温。实际上,水(H_2O)在373K沸腾,在273K凝固,而O_2在90K沸腾,在54K凝固;H_2在20K沸腾,在4K凝固;CH_4在114K沸腾,如此等等。水和冰的结构很难被热所破坏,表明水分子之间有某种非常强烈的连接。

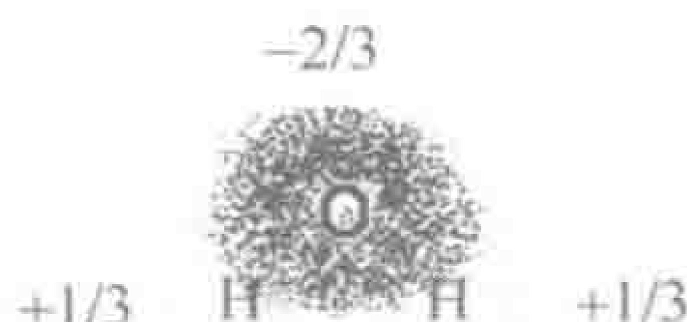
这个“某种连接”,存在于H_2O分子中的O原子和H原子之间(恰恰是在O和H之间,因为单独的O_2和H_2很容易沸腾和熔化)。这种连接被称为**氢键**。

氢键不是只能在水中观察到。只要H原子与一个电负性的(即吸引电子的)原子以共价键相连接的同时又与另一个电负性的原子相接近,就能出现氢键。例如:O—H ::: O,N—H ::: O,N—H ::: N。但是,像C—H基团这样的,却不能形成氢键:因为C原子是电负性不够强的原子。

水作为溶剂的基本特性与其中强烈的氢键有关。

■　水分子中的氢键的本性是电。它只与电子和电荷有关,而与H原子的原子核无关。这可从以下事实直接得知:尽管D(氘)和H的核质量相差一倍,但轻水(H_2O)和重水(D_2O)的沸点、熔点和热能却几乎完全一致。

水分子是有极性的。这意味着,在它的原子里有少量的(“部分”)电荷:在电负性的O原子上有负电荷,在H原子上有正电荷。在这些原子上的电荷和电子云分布如下图所示。



在上图中电子云的密度以点的密度表示,而数字表示这个极性分子中的原子的部分电荷。电量以质子电荷量的分数表示。当然,按这个单位,质子电荷量等于+1,而电子电荷量为-1。水分子整体电量为0。

极化原子上出现电荷的原因是,电负性的O原子把电子云从邻近的H原子上吸

引过来。结果就是 H 原子上出现了一小部分正电荷,而在 O 原子上出现了负电荷。

大家知道,在真空中相距 r 的电荷 q_1 和 q_2 相互作用时的能量为

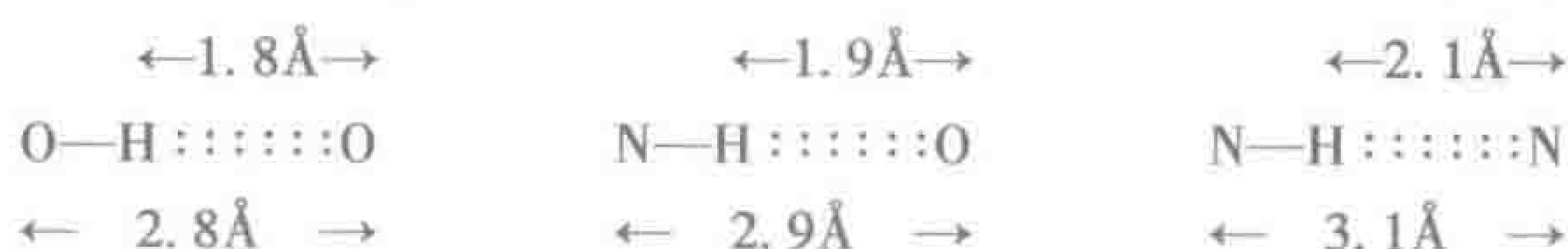
$$U = \frac{q_1 q_2}{r} \tag{4.1}$$

在真空中电荷间的相互作用很强。两个单位电荷(如质子和电子)在相距 1Å 时的相互作用能量大约是 330kcal/mol(请记住这个值:它对我们在后文中的各种估算有用),而在更实际的(考虑到原子的范德华斥力)距离 3Å 时这个能量约为 110kcal/mol。单位电荷的相互作用能量是共价键的特征能量,它比特征热能量 kT 或原子的范德华相互作用特征能量要大 100 倍。

水分子的部分电荷比较小,因此相互作用也比较弱:在距离 3Å 时能量大约为 10kcal/mol,但这么大的能量在 H 原子被吸向 O 原子时已经足以使这两个原子的电子云发生"畸变"。尤其是,电子云的这种"畸变"效应对于氢原子影响极大:H 的电子外壳总共只由单个电子所构成,被吸引靠近 O 原子时,特别容易畸变(补充一句,参与了 H—O 键或 H—N 键的 H 原子的范德华半径只有参与 H—C 键的 H 原子的一半)。而使 O 原子的电子云发生畸变则难得多:围绕 O 原子转动的不仅有 8 个"自己的"电子,还有 O 原子从邻近的氢原子(在 H_2O 中)拉过来的部分电子云。

H 原子的电子云很小而且容易发生畸变,使普通的静电相互作用变成为氢键。在蛋白质中存在的所有氢键都是如此,其中在下文中我们将感兴趣的是:O—H:::O,N—H:::O,N—H:::N。

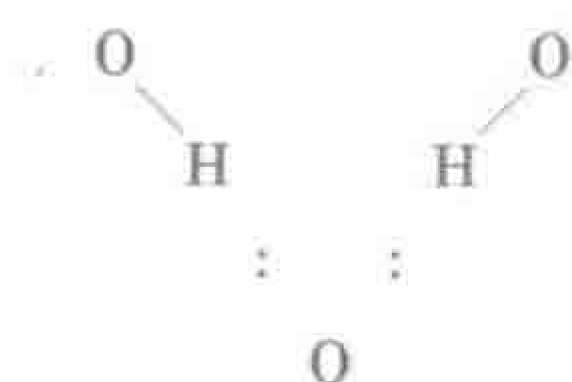
总之,H 原子的电子云最稀薄,在 H 原子所带的部分正电荷与 O(或 N)原子的部分负电荷相互吸引时它会发生显著畸变。结果是 H 原子核与 O(或 N)原子核的距离不是 2.35~2.75Å(我们在上一讲中讨论范德华相互作用时说到的),而是显著地更靠近:相距 1.8~2.1Å(这些数据是从小分子晶体学中获得的)。



这种靠近就产生了**氢键**(或称 H **键**)。H 原子(或更确切地说,是 O—H 或 N—H 基团)称为氢键的**供体**,而该 H 原子所靠近的 O 或 N 原子称为氢键的**受体**。

请注意,在晶体中观察到,形成氢键时供体和受体的 O 和/或 N 原子间的距离大约为 3.0Å(如在冰中为 2.8Å)。这对应于 O 和/或 N 原子间的最佳范德华距离,即它们之间存在的 H 原子并不使供体 O(或 N)原子和受体 O(或 N)原子之间的距离增大:H 原子并不推开它们,而是连接它们。

每个氢键有一个供体和一个受体。H 原子几乎总是只给一个氢键作供体,而 O 原子可以同时作为两个氢键的受体。在后一情况下会出现"叉状"氢键:



由于在形成“叉子”时两个 H 原子(各带有+⅓电荷)非常接近,所以其总能量要小于单个氢键(见前图)能量的两倍。

■　与范德华相互作用不同的是,氢键总是对方向很敏感:尤其是对供体基团的方向。通常,供体(N—H 或 O—H)的共价键直接指向与之形成氢键的受体(O 或 N 原子):



其原因已经涉及 O—H 基团中 H 原子的微小性。氢键对于受体基团的方向不那么敏感:



■　氢键的键能大约是 5kcal/mol。这一估算值是基于对相似化合物(其中一个以氢键结合,另一些不能形成氢键)的汽化热实验值对比得出的。例如,二甲醚($H_3C—O—CH_3$)的汽化热约为 5kcal/mol,而乙醇($CH_3—CH_2—OH$)约为 10kcal/mol。这些物质是由完全同样的原子所组成的(即它们中的范德华作用实际上是相同的),但乙醇能形成氢键,而二甲醚不能(它没有 O—H 基团!),每一个 OH 基团仅可作为一个氢键的供体,而它的 O 原子可以作为氢键的受体。因为每个氢键需要一个供体和一个受体,所以平均每个乙醇分子有一个氢键,于是每个氢键的“能量值” 约为 10kcal/mol(乙醇)-5kcal/mol(二甲醚)= 5kcal/mol。

同样的估算可以根据冰的汽化热获得[680cal/g=680cal/(1/18mol)= 12kcal/mol]。其中有 2kcal/mol 来自范德华相互作用,这可从甲烷(CH_4)或 O_2 之类的小分子的汽化中得出。余下的约 10kcal/mol 是氢键的。在冰中每个 H_2O 分子有两个氢键(图 4-1),因为 H_2O 带有两个 OH 基团,可以作为两个氢键的供体(它的 O 原子也可以接受同样多的氢键)。因此,再一次得到(大约):(10kcal/mol)/2=5kcal/mol。

■　常态冰的结构是由氢键所决定的(图 4-1):它对这些氢键的几何结构(O—H 直接指向 O)是有利的,但对 H_2O 分子间紧密的范德华接触并不那么有利。因此冰结构是透孔的,H_2O 分子在其中包封着微小的(比 H_2O 分子要小)孔洞。冰结构的透孔性导致两种熟知的现象:(1)冰的密度比水要小,它浮在水上;(2)在很大的压力下(如冰刀)冰会融化。

存在于冰(图 4-1)中的氢键绝大部分继续存在于液态水中。这可从冰的融化热(80cal/g)远小于水的汽化热(0℃时为 600cal/g,100℃时为 540cal/g)而推导出来。可能会认为,在冰中存在的氢键,在液态水中仅有 80/(600+80)= 12% 断裂了。但是这种图景(部分氢键在水中断裂,部分保留)并不准确:确切地说,是水中所有氢键都松弛了。下述实验数据能很好地说明这一点。

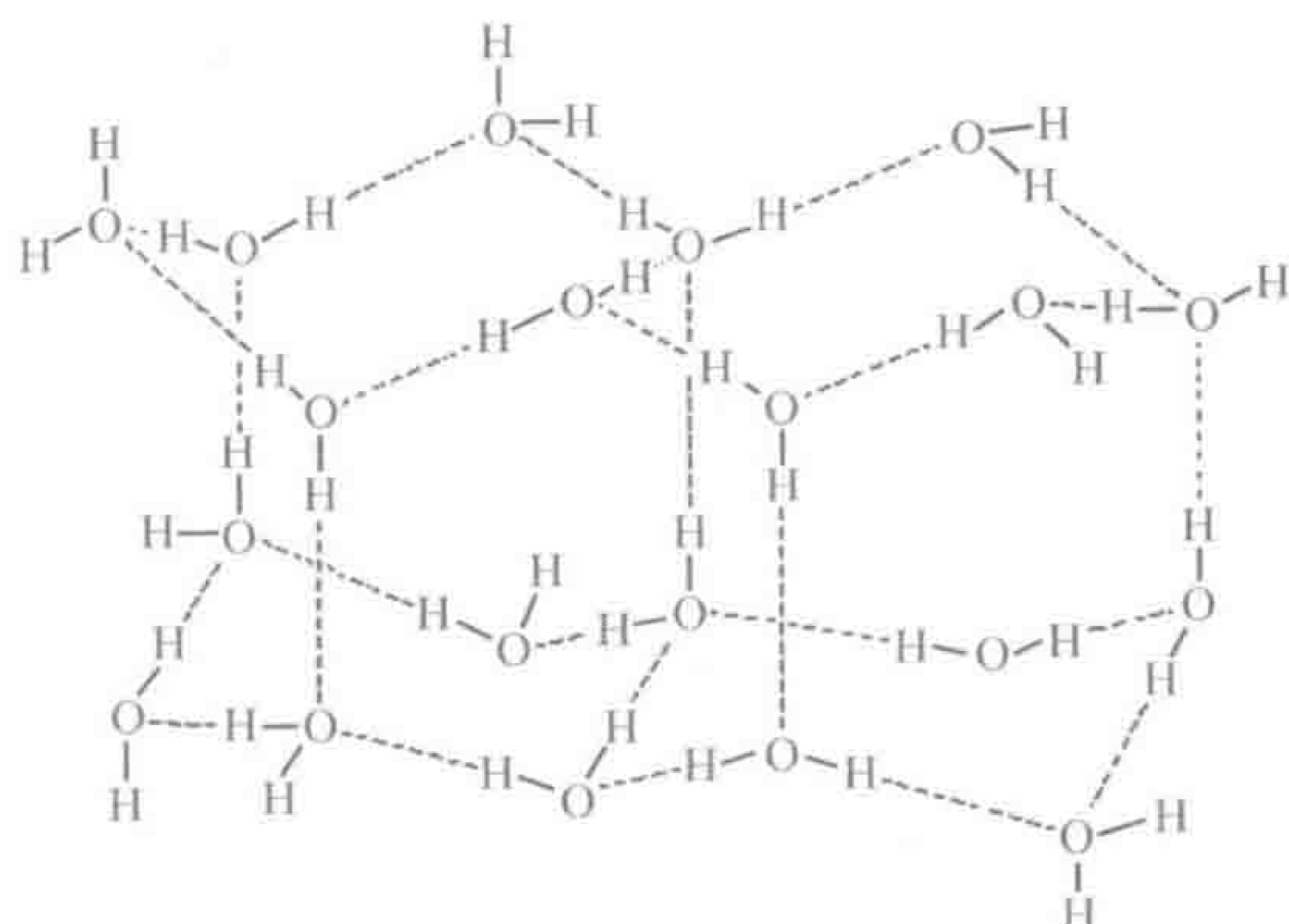


图 4-1　常态冰（“I_h 冰”；还有其他形式的冰，但只在极高压强下才稳定）。虚线表示氢键。在冰的透孔结构中可见的由 H_2O 分子包围的小空腔。图片取自文献[6]，略有改动。在常态冰中每一个 H 原子仅与一个 O 原子以共价键结合。实际上，正如 X 射线分析所显示的，每一个 H 原子可以以相等的概率处于两种状态：要么位于与它以共价键相连的 O 原子（图中实线连接）旁边，要么位于与它以氢键相连的 O 原子（图中虚线连接）旁边。H 原子从一个 O 原子旁转换到另一个 O 原子旁会引起这两个 O 原子以及在它们周围的 O 原子的共价键和氢键的重新构建。结果是在冰中的整个共价键和氢键网络强烈地振荡，而且这导致冰的不同寻常的高极化性：它的介电系数在 0℃ 时为 97，超过液态水的介电系数（在 0℃ 时为 88）

在图 4-2 中，曲线 1 为冰中（此时所有氢键都饱和）的 O—H 基团的红外吸收谱的最大值；曲线 2 为溶于 CCl_4 中的独立 H_2O 分子（此时不存在氢键，H_2O 在 CCl_4 溶液被极大地冲稀了）的 O—H 基团的红外吸收谱；而曲线 3 为液态水的吸收谱。如果在水中存在两种 O—H 基团（形成氢键的和未形成氢键的），则水中的一种 O—H 基团会像冰中的那些 O—H 基团（全部形成氢键）那样振动（频率相同），而另一种则像 CCl_4 溶液中的（未形成氢键）那样。那么水的吸收光谱会具有对应于 O—H 基团两个状态、两个特征振动频率的两个最大值：相同的基团振动频率会吸收相同频率的光。但是并**没有**观察到“双峰”曲线！在曲线 3 上我们观察到一个平缓的最大值，从曲线 1 的最大值伸展到曲线 2 的最大值。这意味着，在液态水中所有 O—H 基团由氢键所连接，但是几乎所有这些氢键都松弛了，而且松弛程度各不相同。

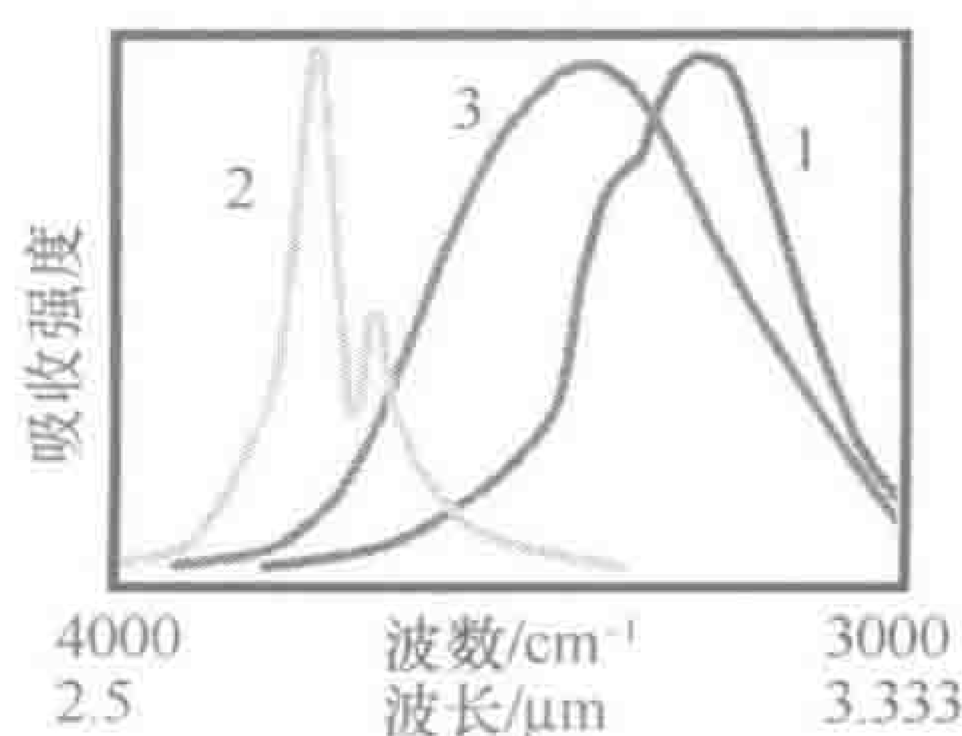


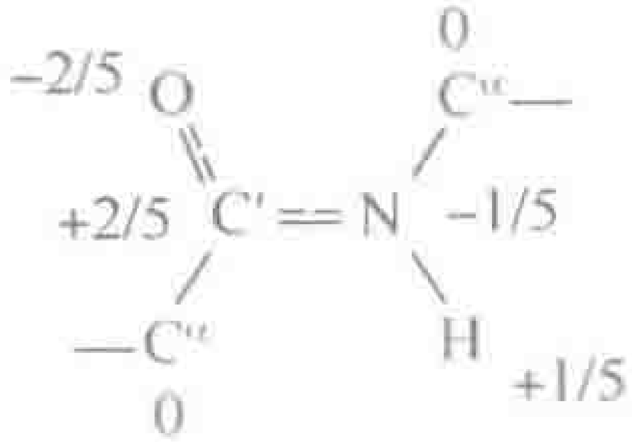
图 4-2　在冰中（1），在 CCl_4 环境中（2）和液态水（3）中的 O—H 基团的远红外吸收谱的典型形状。根据 Greinacher E.，Luttke W.，Mecke R.1995.Electrochem Z.59：53；Золотара.В.М. 1970.Onmuka u cnekmpockonuя.29：1125 绘制

这表明，“在水中，部分氢键断裂，部分氢键保留”的说法，严格地说，是不正确的。但是它在描述水的热力学性质时很简单、很方便，所以被广泛地使用着，而且我们也

会使用它。但是心里应该记得,这是不准确的,在使用这个模型时需要小心。

■ 现在来看蛋白质链与水的相互作用。

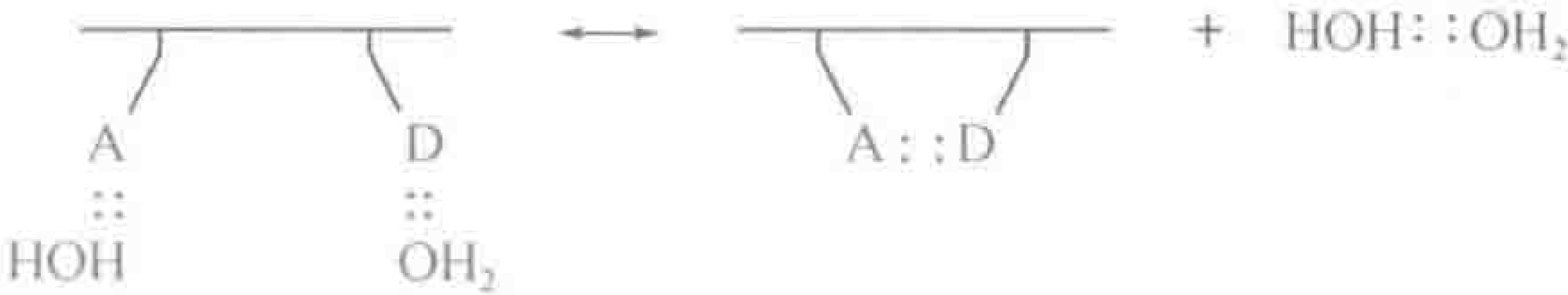
蛋白质链的骨架与水分子一样,是有极性的。更确切地说,位于骨架内的肽基团是有极性的。在总体上蛋白质链的骨架净带电量为 0,但是电荷在它的原子上(这里每个原子上的电量也是以质子电荷量为单位)的分布如下所示:



有几个侧链基团也带有部分电荷,如丝氨酸(Ser)(它的侧链为—CH_2—OH,与乙醇相似)。带电的氨基酸残基极性化程度更大:当残基精氨酸(Arg)、赖氨酸(Lys)和组氨酸(His)处于离子化形式时(大约是中性 pH),它们的碱性侧链基团带有电量 +1,而天冬氨酸(Asp)和谷氨酸(Glu)的离子化酸性侧链基团所带电量等于-1。

主链的肽基团和侧链的极性基团都可作为氢键的供体和受体。它们可以相互之间形成氢键,或与水分子形成氢键:实际上它们全部都以氢键相连接,由于氢键能量(5kcal/mol)比热运动能量高出 10 倍,故热运动不能全部破坏这些氢键,只是时不时地偶尔破坏其中一小部分。

在水环境中,如果在氢键供体(D)和受体(A)之间形成蛋白质分子内的连接,则此连接替换了蛋白质与水分子形成的两个氢键(它们原先存在着,因为不能有空位);同时还会在两个脱离开的水分子之间形成另一个氢键。



这种反应的能量变化约等于零:由两个氢键得到两个氢键。但是水的**熵**增加了,因为水不再与蛋白质链相连,而且由氢键相连的分子可以游到任意地方去(而熵正是与系统状态的数目多少相关)。在这种情况下水二倍体脱离蛋白质而产生的熵增量大约与 H_2O 分子脱离冰进入液态水的熵增量相等(在两种情形下都是一个粒子获得运动的自由)。

H_2O 由冰进入水中的熵效应可以由下述方法估算。已知,在熔点(冰是 0℃,即 273 K)融化时的熵增量刚好能补偿相应的能量损失。而这种损失我们已经知道:80cal/g,即 80cal/g×18g/mol =1440cal/mol(18g/mol 为 H_2O 的摩尔质量)。

因此,由于水的熵增加,"水中的蛋白质"系统的自由能降低了:每形成 1 摩尔分子内的氢键 D:::A,其自由能减小约 1.5kcal。

水释放的这些自由能,可以全部或部分地补偿由形成 D:::A 键而使 D 和 A 基团固定(原先它们像尾巴一样摆动,而现在结合后不能摆动了)时所伴随的熵减少而增加的

自由能。后面我们会看到，水自由能的这种减少大致能补偿固化多肽链二级结构中氢键单体（N—H∷∶O）的熵。结果是水中多肽的规则二级结构刚好处于稳定性的边缘。

■　于是，奇怪的名词："熵"和"自由能"被说出来了。我的经验表明，普通的生物学学生听说过"熵"（他们听说，这是状态数目多少的度量，无序度的度量），而"自由能"完全不知道是什么……由于我们经常需要应用"自由能"的概念，我想现在在这里花几分钟讲一下，然后在必需的时候再更详细地讨论它（以及熵）。

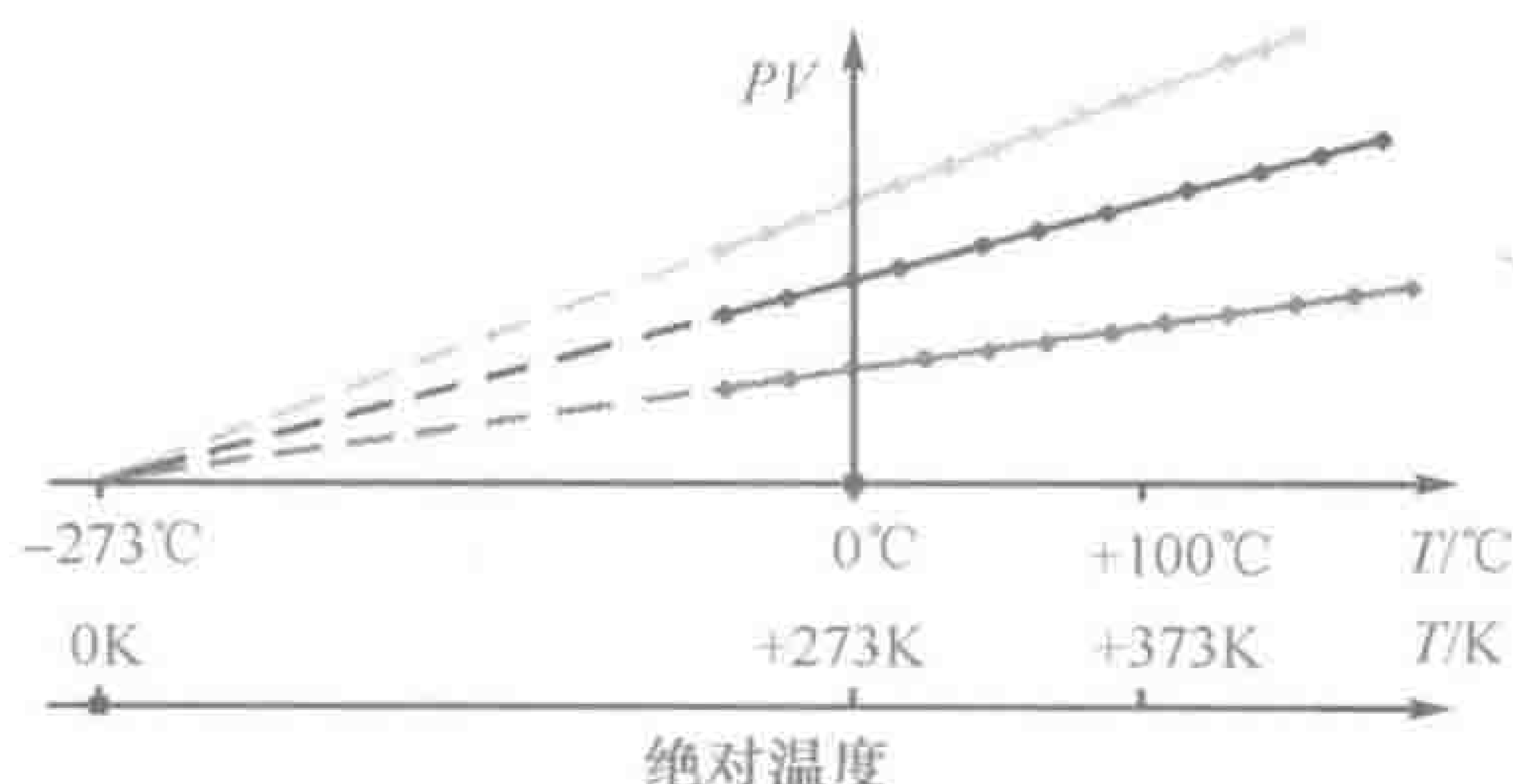
以一个简单的例子开始吧：假设分子可以处于两种状态，"a"和"b"，比如，在这里，在这个房间里，海拔 200m 的地方（"a"），和在喜马拉雅山上的布达拉宫里，海拔 5km 处（"b"）。假设：(1) 两地的温度 T 都相同，(2) 我们有足够长的时间来观察我们的分子直到它们能达到这两个地方，那么这两种状态的概率之间的关系如何？

通常所有人都知道玻尔兹曼公式，内容如下：

处于能量为 E 的状态的概率　正比于 $e^{\frac{-E}{kT}}$　　(4.2)

该公式的物理意义在于，外界环境的热能（与其他分子的碰撞）能以某种程度（平均地，正比于环境温度 T）激活我们的分子，从而使它到达能量稍微高一点的地方。对此我们将会在后面更详细地讨论，而现在我冒昧地假设，你们已经理解了这个公式。还提醒一下，k 是玻尔兹曼常量，T 为绝对温度，以开尔文温标表示（它从"绝对零度"开始计数，0K = −273.16℃）.

不，让我们还是不要指望大家都知道这些吧。让我们从绝对温度开始。为什么从克拉贝龙和开尔文时代开始，就从−273℃开始读数？因为这个点是唯一的（这与0℃不同：0℃是千万种物质之一水的凝点）。由此外推，如下图所示，**任何**数量的**任何**稀薄气体的压强与体积的乘积（$P\times V$）在接近低温时，可见在−273℃时趋向于0。



现在我们来推导玻尔兹曼公式，至少就我们感兴趣的最简单的例子，即气体分子随高度的分布来推导吧。此时能量 $E(h)=mgh$，其中 m 为气体分子质量，g 为重力加速度，h 为高度。

已知（记得克拉贝龙-门捷列夫公式①吗？一分钟前才用它确定绝对温度），理想气体的压强 P 等于 nkT，其中 n 为单位体积内的气体分子数。如果 T 不随高度 h 而变化，则 $\frac{\mathrm{d}P}{\mathrm{d}h}=\frac{\mathrm{d}n}{\mathrm{d}h}kT$。另

①国内教科书一般称为"克拉贝龙方程"，即理想气体状态方程 $PV=nRT$，n 为体积 V 中所含的分子摩尔数。

一方面，考虑单位横截面上的气体柱，我们看到，由于每上升 dh，压在该单位横截面上的气体分子总重量减少(mgn)dh，有 $\mathrm{d}P=(mgn)(-\mathrm{d}h)$。因此 $\frac{\mathrm{d}P}{\mathrm{d}h}=\frac{\mathrm{d}n}{\mathrm{d}h}kT=-mgn$。所以，$\frac{\mathrm{d}n}{\mathrm{d}h}=-\frac{mg}{kT}n$，或 $\frac{\mathrm{d}\ln n}{\mathrm{d}h}=-\frac{mg}{kT}$，即 $n\sim \mathrm{e}^{-mgh/kT}=\mathrm{e}^{-E(h)/kT}$。

于是，应用到我们感兴趣的问题：分子停留在不同高度的概率之比如何，玻尔兹曼公式导出了**气压关系**：

$$\frac{\text{处在高度 } b \text{ 的概率}}{\text{处在高度 } a \text{ 的概率}}=\frac{\mathrm{e}^{-E_b/kT}}{\mathrm{e}^{-E_a/kT}} \tag{4.3}$$

式中，E_a 为分子在状态"a"(即"这里")的能量；E_b 为分子在状态"b"("5km 高处")的能量；而 T 为绝对温度(如前所述，为简单起见假设它不随高度变化)。

由于重力作用，分子在"这里"的能量比"5km 高处"要低，因此，根据玻尔兹曼公式，分子呆在"在 5km 高处"的时间将会比"这里"要短。

[减少至 1/1.5~1/2；大家可以在 $T=300$K 自己计算，提醒一下，$E=mgh$，其中 m 为空气分子的平均质量($\approx$30 道尔顿 = 30 g/mol)，$g\approx 10\mathrm{m/s^2}$，为重力加速度，$h\approx$5km，为高度差，而玻尔兹曼常量 k 等于每粒子 1.38×10^{-23} J/K(1J = 1kg · m^2 = 0.24cal)，即每摩尔(6×10^{23})粒子 2cal/k。提醒一下，R=2cal/(mol · K)，为理想气体常数。]

这说明在 5km 高处的分子数只有这里的 1/2。更确切地说，是**在相同的单位体积**中。例如，在我们的肺中(对此只要试试在不同高度呼吸就很容易信服)。但是，在布达拉宫中**总的**分子数比这个房间里要多：因为宫殿的空间比我们这个房间要大。即那里的分子可以处于数目更多的位置：对于自由飞行的空气分子而言，位置的多少正比于房间的体积。在这种情形下物理学家们会说，分子在该宫殿中的**微观状态**数比在我们这个房间中要多。因此分子处于宫殿中**某处**的概率，与处于我们这个房间中某处的概率之比为

$$\frac{\text{处于体积"}b\text{" 中某处的概率}}{\text{处于体积"}a\text{" 中某处的概率}}=\frac{V_b\mathrm{e}^{-E_b/kT}}{V_a\mathrm{e}^{-E_a/kT}}$$

式中，V_a 为状态"a"("我们这个房间")的体积；而 V_b 为状态"b"("布达拉宫")的体积。回想一下大学基础数学，可知 V 可以表示为 $\mathrm{e}^{\ln V}$，则上式可改写为

$$\frac{\text{处于体积"}b\text{" 中某处的概率}}{\text{处于体积"}a\text{" 中某处的概率}}=\frac{\mathrm{e}^{-E_b/kT+\ln V_b}}{\mathrm{e}^{-E_a/kT+\ln V_a}}=\frac{\mathrm{e}^{-(E_b-kT\ln V_b)/kT}}{\mathrm{e}^{-(E_a-kT\ln V_a)/kT}}$$

上面这个式子很像式(4.3)，即玻尔兹曼公式，但它不是应用于**单位体积**，而是应用于**整个系统的体积**，并且(注意!)在能量 E 的位置写着 $E-T\times k\ln V$。正是 $F=E-T\times k\ln V$ 这个值被称为空气分子在给定体积 V 和温度 T 时的**自由能**。而值 $S=k\ln V$ 被称为空气分子在体积 V 中的**熵**。

具体到本例中，"分子可及状态数"就是分子可及体积 V(V_a 为"我们这个房间"，V_b 为"布达拉宫")。

在一般情形下，熵 S 简单地等于 $k\times$［**可及状态数的对数值**］。而自由能 F 与能量 E、熵 S 和温度 T 的关系如下式所示：

$$F = E-TS \tag{4.4}$$

在两种状态中，更稳定的，即**概率更大的**是自由能较低的那一个：

$$\frac{\text{处于体积“}b\text{”中某处的概率}}{\text{处于体积“}a\text{”中某处的概率}}=\frac{e^{-F_b/kT}}{e^{-F_a/kT}}=e^{-(F_b-F_a)/kT} \tag{4.5}$$

■ 换句话说，概率更大的、**更稳定的**系统状态对应于自由能 F 较低的状态。而**最**稳定的系统状态（在给定温度和该系统的体积时）对应着**自由能 F 最小值**。

因此，“自由能”是普通的“能量”概念在**系统与外界环境交换热量**时的自然扩展。记住：如果物体与环境**不**交换热量，它的稳定状态对应着它的能量最小值（或更简单地说：所有能落下的东西，最终都会落下）。而当系统的分子被外界热量所激发时，它们开始获得无数具有更高能量的状态（即系统的熵增加），结果就是，空气分子飞起来，而不落在地面上。

■ 或者这样说：

能量的变化 E_b-E_a，是在**物体与外界环境没有热交换**时，把一个物体从状态“a”转变到状态“b”所需要做的功。

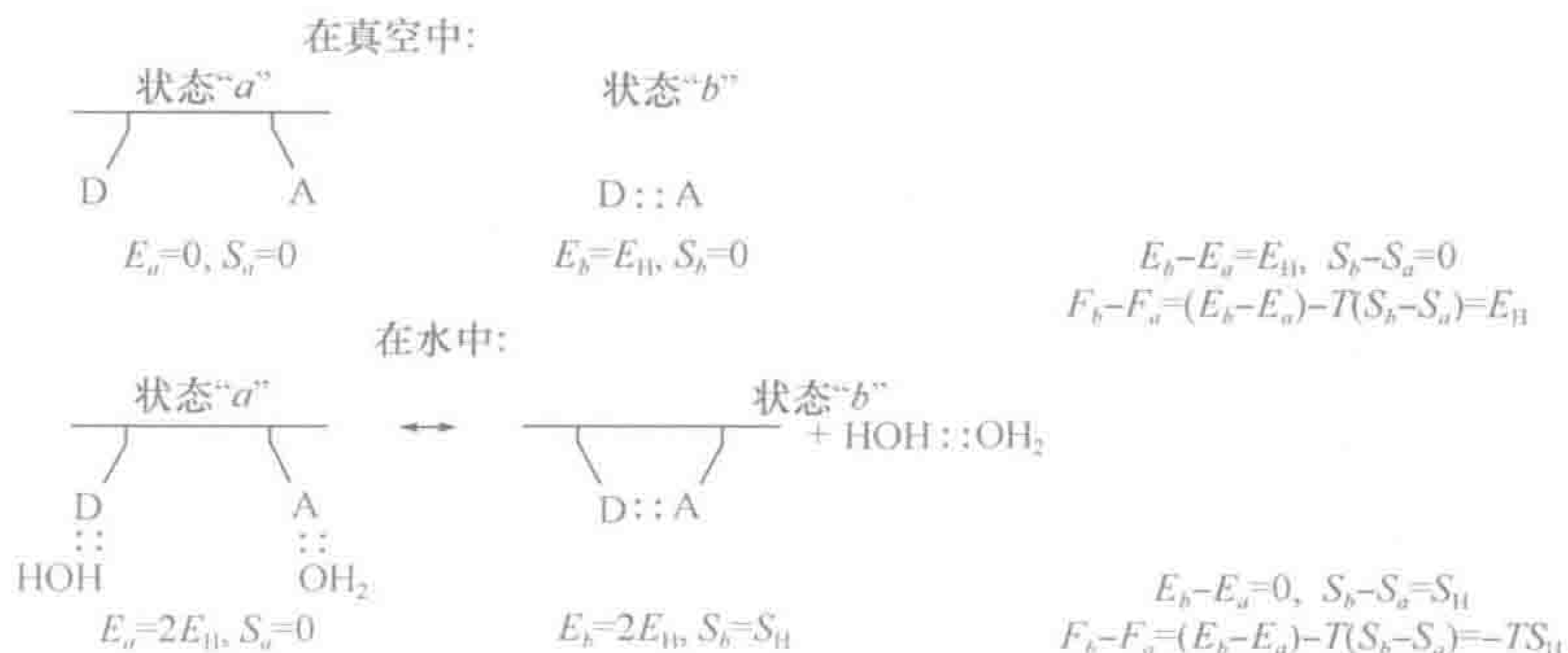
而自由能的变化 F_b-F_a，是在**物体与外界环境保持热交换**时，把一个物体从状态“a”转变到状态“b”所需要做的功。

■ 让我们从喜马拉雅山回到蛋白质上来。

那么，在前面我们讨论的在蛋白质链中形成氢键的例子中，能量、熵和自由能之间的平衡是什么样的？

为了完全直观起见，让我们来比较在不同条件下的进行过程：

这里氢键的能量 $E_H<0$，而自由物体（如游离 H_2O 分子或游离的 HOH:OH_2 二倍体）的平动和转动熵 $S_H>0$。当 $E_H-TS_H<0$ 时，氢键（水与水之间的，或水与蛋白质之间的）是稳定的（若 $E_H-TS_H>0$，即它们不稳定，则在这种情形下我们面对的已经不是液态水，而是水蒸气）。



上面示意图中的比较表明,由于结合水的熵变化,在水中的蛋白质链中的氢键变得不如真空中那么稳定;实际上,在水中它们的自由能为 $F_b - F_a = -TS_H$,即其绝对值比在真空中($F_b - F_a = E_H$)要小。

我想再次强调,这种氢键明显减弱的原因在于,在水环境中蛋白质链中形成的氢键**取代**了蛋白质链与水之间的氢键。基于同样的原因,使**水中的**蛋白质结构更稳定的氢键具有熵性质,而不是能量性质:链的两种状态(有、无内部氢键)的能量大致相等,而具有大致相等能量的这两种状态中,更稳定的是熵更高,即微观状态数更多的那一种。而被释放的水分子的微观状态数比被束缚的水分子的微观状态数要多。

注意:**在水中的**蛋白质链中的氢键具有熵性质而**不是**能量性质,**原因**在于氢键的能量非常高:正因为如此,在链上“未参与”**蛋白质中**的氢键的氢键供体和受体实际上并非不参与**任何**氢键,而总是与水以氢键结合。在形成蛋白质内部氢键时,从蛋白质释放的 H_2O 分子很快就互相结合,因而能量得以被补偿,而蛋白质中氢键的自由能增益仅仅来自于所释放的水分子的可能微观状态数的增加。确实,H_2O 分子为了相互结合而必须牺牲掉一部分所获得的自由度、部分熵,但是,失去少量的熵比失去大量能量更好!

有两个原因决定了水作为溶剂的特殊性:①水分子相互之间以氢键紧密连接;②这些氢键只有在水分子之间的特定相互取向上形成。这导致了种种有趣的效应。对此我们将在下面两讲中详细讲述。

第 5 讲(两倍*)

热力学基础。自由能与化学势。疏水作用。它们与氢键在水中的饱和必然性的联系。氨基酸的水可及非极性表面及其疏水性。

在本讲的前面部分我想继续介绍热力学。这对我们以后讨论水作为溶剂的特殊性是有益的。在这些讨论中,各种分子浸没于水中的自由能扮演着关键的作用。

■ 为了研究分子浸没于水中的自由能,我们取一个封闭的试管,一半装水,另一半装水蒸气、空气或有机溶剂,并观察所研究的分子在这两个相之间如何分布。

我们已经看到,根据下式,自由能 F 之差决定所研究系统的最有利状态(当前情况是,所研究的粒子的最有利位置):

$$\frac{\text{处于“}b\text{”中某处的概率}}{\text{处于“}a\text{”中某处的概率}} = e^{-(F_b - F_a)/kT} \tag{5.1}$$

自由能 $F = E - TS$ 由系统能量 E 及其熵 S 所组成。对于能量,想必所有人都理解。对于温度,更主要是凭直觉理解,我们还会回头来讨论。至于熵,更复杂一些,让我们再来讨论一下。

对于粒子处于容器中的最简单情形,如果你们记得,我们已经讲过了,$S = k\ln V$,其中 V 为粒子可及体积。

这个熵进入自由能(以 $-TS$ 的方式)有什么意义?如果我们丢掉 E,来看 $e^{-F/kT}$,则有 $e^{-(-TS)/kT} = e^{-(-Tk\ln V/kT)} = V$,就是可及空间,而它决定了粒子在空间中的**可能状态数**。熵越高,这个状态数越多,因而粒子处于这个空间中某处的概率越大。

在一般情况下(当粒子的可能状态不仅受限于它与环境体积 V 的侧壁的碰撞,而且受限于它与其他粒子的碰撞),粒子的熵 S 等于

$$S = k \times [\text{所研究粒子的可及状态数的对数值}]$$

在分子物理学、生物学、化学中通常讨论的不是单个粒子,而是 1mol(6×10^{23})粒子,于是写为

$$S = R \times [\text{单个粒子的可及状态数的对数值}]$$

这里需要再次提醒,k 与 R 之间的差别仅仅在于,在讨论单个粒子时使用 k,而讨论 1mol 粒子时使用 R。

严格地说,物理量

$$F = E - TS \tag{5.2}$$

*本章内容较多,是前面章节的两倍。

被称为**亥姆霍兹**(Helmholtz)**自由能**。它容易描述,计算方便,因为这个数值说明的是处于封闭的固定体积中的系统(正与我们讲过的分子一样)。

但是通常在实验中测量不是在恒定体积 V 中进行,而是在恒定内部压强 P(如大气压强)下进行。在这种情况下,测量的不是所研究物质的粒子能量 E 的变化,而是它的**焓**$H=E+PV$ 的变化:在焓 H 的变化中,实际上除了物体的能量(E)变化之外,还包含这个物体的体积 V 改变时反抗恒定内部压强 P 所做的功。

在实际中(在本课程中我们所讨论的就是这类情形),数值 PV 可以忽略不计,因为我们将感兴趣的将是液态或固态物体(这时每个分子处于很小的体积中),而且是在不高的压强下(大气压强级别)。在这些条件下,数值 PV 远小于此物体的热能。

实际上,甚至对于气体(此时每个分子的体积都相当大),$PV\approx RT$([物质的量](记得克拉贝龙-门捷列夫定律吗?)。因此,在计算 1mol 物质时,在温度为 0~100℃,即 T=273~373K 时,修正值 $PV=1RT\approx0.55\sim0.75$kcal/mol。因此甚至在气体中,它与我们感兴趣的效应(通常为几个 kcal/mol)相比也是不大的。对于液体和固体(在大气压强级别时),PV 的修正值在 H 值中要小得多:此时在 1atm①、室温下 1mol 物质不像气体那样占据 25L,而只是 1L 的很小部分(水是约 1/55L,$(CH_2)_6$是 1/10L,如此等等)。

因此,在后文中 H 与 E 的差别对我们并不重要。对这两个值我将通常简称为"能量"。

类似地,亥姆霍兹自由能 $F=E-TS$ 与**吉布斯自由能**

$$G=H-TS=(E+PV)-TS=F+PV \tag{5.3}$$

之差对我们也并不重要。因此对这两者我们也不作区别,将会简单地称为"自由能"。

尽管如此,需要记住,讨论在封闭体积中进行的过程时必须使用符号 E 和 F,而讨论在恒定内部压强中进行的过程时必须使用符号 H 和 G。

还有,把新粒子添加到系统中所需要做的功被称为**化学势**。稍迟一点我们会讨论它。

一点注释。如果在恒定体积中向系统(如液体)中逐个添加粒子,则添加新粒子时所需做的功将随着系统中的粒子数增加而改变(向封闭的瓶子里注射 1 滴并不困难;但是第 2 滴呢? 第 3 滴呢……)。但是如果在**恒定压强**下这样做,则添加一个新粒子所需做的功并不随着系统中的粒子数而改变,即在恒定体积下 E 和 F 并不正比于系统中的粒子数,而 H 和 G (在恒定压强下)与之成正比。后者,如果系统中相同粒子数很多时,能以很方便的形式确定化学势:"单个粒子的吉布斯自由能"。

需要记住,在给定温度下(T=const)任意系统**如果**在固定的体积内达成平衡,则系统的平衡、稳定状态对应着数值 $F=E-TS$ 的最小值,**如果**是在恒定的内部压强下达成平衡,则对应着 $G=H-TS$ 的最小值。

---

①1atm=101 325Pa,下同。

系统的状态发生微小变化时，它的自由能发生轻微变化：

$$F \to F + \mathrm{d}F = \mathrm{d}E - T\mathrm{d}S - S\mathrm{d}T$$
$$\text{或 } G \to G + \mathrm{d}G = \mathrm{d}H - T\mathrm{d}S - S\mathrm{d}T \tag{5.4}$$

这意味着，尤其是系统在其稳定状态附近的所有可能的重建（它们可能会改变系统的能量和熵）遵从下述描述函数最小值的公式（正如我们所知，函数最小值具有如下性质：对它的微小偏离几乎并不改变这个函数的值）。

1）在 $V=\text{const}$ 时，稳定状态（给定 T 时）对应着 F 最小值，即

$$\mathrm{d}F\,|_{V=\text{const}} = \mathrm{d}E\,|_{V=\text{const}} - T\mathrm{d}S\,|_{V=\text{const}} = 0 \tag{5.5}$$

［这里我们已知，在 $T=\text{const}$ 时有 $\mathrm{d}T=0$，即式(5.4)中有 $S\mathrm{d}T=0$］；

2）在 $P=\text{const}$ 时，稳定状态（给定 T 时）对应着 G 的最小值：

$$\mathrm{d}G\,|_{P=\text{const}} = \mathrm{d}H\,|_{P=\text{const}} - T\mathrm{d}S\,|_{P=\text{const}} = 0 \tag{5.6}$$

从这些式公式推导出绝对温度的热力学定义：

$$T = \left.\frac{\mathrm{d}E}{\mathrm{d}S}\right|_{V=\text{const}} = \left.\frac{\mathrm{d}H}{\mathrm{d}S}\right|_{P=\text{const}} \tag{5.7}$$

我知道，现在这个定义是相当形式化地推导所得，它的物理意义不大明显。因此我将会在下一次课中回到这个问题上来。

■　现在来看**化学势**：该值决定的不是系统整体的热力学性质，而是该系统中的单个分子的热力学性质。

如果在**恒定压强**下逐个向系统添加分子，则在添加每个新粒子时需要花费与添加任意前一个粒子时完全相同的功：系统体积将会增加，而系统的密度（以及其中的相互作用强度）不会改变。因此在均质大系统中的分子的热力学状态很方便地可由吉布斯自由能 G 除以分子数 N 而确定：

$$\mu = G/N \tag{5.8}$$

μ 被称为**化学势**（因为在压强不大的液相或固相中 $F \approx G$，所以此时有 $\mu \approx F/N$）。（如果 N 不是表示分子数，而是，正如通常在物理化学中的那样，表示分子的摩尔数，则 μ 不是属于单个分子，而是属于 1mol 分子）

计算单个分子的化学势以及吉布斯自由能，在本讲中的第二部分，即讨论分子在相间的分布时，对我们是有益的。关键在于，分子从它的化学势较高的相流到它的化学势较低的相，这会降低系统总的自由能并使它接近平衡。而在平衡时分子在一个相中的化学势等于同类分子在另一个相中的化学势。

■　我们在后面还需要两个公式。

首先是热容的确定，该量表明系统能量随温度的增加：

$$C_P = \left.\frac{\mathrm{d}H}{\mathrm{d}T}\right|_{P=\text{const}} \tag{5.9}$$

这是在恒定压强下的；也可以在恒定体积下计算热容，但我们并不需要这些细节。

第二，熵与自由能的关系：

$$S=-\left.\frac{\mathrm{d}G}{\mathrm{d}T}\right|_{P=\mathrm{const}}=-\left.\frac{\mathrm{d}F}{\mathrm{d}T}\right|_{V=\mathrm{const}} \qquad (5.10)$$

式 (5.10) 是最重要的热力学公式。它是这样直接推导而得：自由能的微小增量 $\mathrm{d}G=\mathrm{d}(H-TS)=\mathrm{d}H-T\mathrm{d}S-S\mathrm{d}T$ (且 $\mathrm{d}F=\mathrm{d}(E-TS)=\mathrm{d}E-T\mathrm{d}S-S\mathrm{d}T$)，同时(在系统平衡下)根据**温度的热力学定义**$T=\mathrm{d}H/\mathrm{d}S=\mathrm{d}E/dS$ [式(5.7)]有 $\mathrm{d}H-T\mathrm{d}S=0$ (且 $\mathrm{d}E-T\mathrm{d}S=0$)。式 (5.10) 表明，自由能在满足 $S(T)=0$ 的温度 T 下达到自己的最小(或最大)值。

同样，知道这一点也是有益的：G/RT 在 $H(T)=0$ 时达到自己的最小(或最大)值，因为有：

$$\frac{\mathrm{d}\left(\dfrac{G}{RT}\right)}{\mathrm{d}T}=\frac{\dfrac{\mathrm{d}G}{\mathrm{d}T}}{RT}-\frac{G}{RT^2}=-\frac{S}{RT}-\frac{H-TS}{RT^2}=-\frac{H}{RT^2} \qquad (5.11)$$

注释：

1. 由式 (5.3)和式(5.10) 可推出：$H=G+TS=G-T(\mathrm{d}G/\mathrm{d}T)$，因而[式 (5.9)]有

$$C_P=-G\left.\frac{\mathrm{d}^2G}{\mathrm{d}T^2}\right|_{P=\mathrm{const}} \qquad (5.12)$$

2. 我们对能量、熵或自由能本身的数值从来都不感兴趣；我们感兴趣的是这些量的**变化**。确实，重力势能可以从海平面、地面或地心等开始计算，结果将会各不相同，但这并不重要，重要的只是粒子在不同位置时这些能量值的**差值**。同样的，确定了粒子的熵为 $k\ln V$ 之后我们可以用 L、m^3 或其他单位来量度体积 V，这个熵的数值也会不同，但这并不重要，重要的只是粒子在两个体积(V 和 V')中的熵值之**差**。而这个差值，$\Delta S=k\ln(V/V')$，与体积的测量单位无关。

在这些开场白之后我们准备好了讲述水作为溶剂的特殊性。首先来看所谓**疏水效应**。

■ “疏水(Hydro-phobicity)”的意思就是“怕水”。谁怕水？所有非极性分子：惰性气体(氩气、氙气)，氢气，以及所有纯烃类分子(甲烷、乙烷、苯、环己烷等)。我们所关注的正是烃类在水中的性质：因为在蛋白质中有很多氨基酸带有烃类侧链基团。恰恰是这些基团害怕水并远离水，从而形成**蛋白质球体的疏水核**。

那么，在实验上疏水意味着什么？如果我们改变甲烷(CH_4)在水中和水上方的水蒸气中的浓度，则在水蒸气中的甲烷浓度将**多**出 10 倍(准确的数值与温度有关，不过在室温下 20~40℃时，大约是 10 倍)。对 H_2 和丙烷 $CH_3CH_2CH_3$ 也有相近的比例关系：所有它们都害怕(疏远)与水混合，它们在水蒸气中都比在水中要**多**。而乙醇 CH_3CH_2OH 能与水很好地混合，要分离它们相当困难——对此大家都知道；但乙醇是极性分子(它的极性 O—H 基团能结合成氢键)。而纯**非极性**分子更愿意处于真空中(水蒸气几乎是真空)，而不是水中……**虽然任意**分子(包括 H_2 和 CH_4)对 H_2O 都有范德华吸引力！

来看一个烃溶于水的热力学效应的典型例子(图 5-1)。图 5-1 中的热力学数据(在相间迁移的自由能、能量和熵)是基于对环己烷(CH_2)$_6$ 分子在水蒸气、水溶

剂和液态环已烷这三相之间的平衡分布的研究而得到的。

分子从非极性液体迁移到水中的自由能由下式决定：

$$\Delta G_{\text{非极性液体}\rightarrow\text{水}} = -RT\ \ln(X_{\text{水}}/X_{\text{非极性液体}}) \tag{5.13}$$

这里 $X_{\text{水}}$ 和 $X_{\text{非极性液体}}$ 为所研究的分子在水溶液中和与水溶液接触的非极性液体中的平衡浓度。$\Delta G_{\text{非极性液体}\rightarrow\text{水}}$ 为在水中和非极性液体中作用于我们的分子上的“平均力势”之间的差值。平均力势是由分子与其周围环境之间的所有相互作用力建立起来的；它包括由这些相互作用产生的能量成分和熵成分。由于浓度是给定体积中的分子数，故式(5.13)对应着平均力势不同的两相之间的玻尔兹曼分布。

纯液态 $(CH_2)_6$ 在 25℃ 和低压强(~1atm)下，$X_{\text{非极性液体}} = 9.25$mol/L，对其饱和水溶液，$X_{\text{水}} = 0.0001$mol/L。因此，$\Delta G_{\text{极性液体}\rightarrow\text{水}} = +6.7$kcal/mol。

分子从气体迁移到非极性液体或水中的自由能也由同样的方式确定：

$$\Delta G_{\text{气}\rightarrow\text{非极性液体}} = -RT\ \ln(X_{\text{非极性液体}}/X_{\text{气体}}) \tag{5.14}$$

$$\Delta G_{\text{气}\rightarrow\text{水}} = -RT\ \ln(X_{\text{水}}/X_{\text{气体}}) \tag{5.15}$$

式中，$X_{\text{气体}}$ 为分子在与非极性液体或水相接触的气体中的平衡浓度。饱和 $(CH_2)_6$ 蒸汽在 25℃时的压强约为 0.05 atm，对应着 $X_{\text{气体}} = 0.002$mol/L(即 $X_{\text{气体}}$ 比 $X_{\text{水}}$ 高出 20 倍)。相应的 ΔG 值显示于图 5-1 中。

由于在稀薄气体($X_{\text{气体}}$)中作用于分子上的平均力势实际上等于零，故 $\Delta G_{\text{气体}\rightarrow\text{非极性液体}} = G_{\text{非极性液体}} - G_{\text{气体}}$ 的值非常接近于 $G_{\text{非极性液体}}$，即在非极性液体中作用于分子上的平均力势；而在水中作用于分子上的平均力势 $G_{\text{水}}$ 实际上等于 $\Delta G_{\text{气}\rightarrow\text{水}}$。

若已知 ΔG 值及它随时间的变化，就可以根据式(5.10)、式(5.11)和式(5.12)求出 ΔS、ΔH 和 ΔC_p 的值。

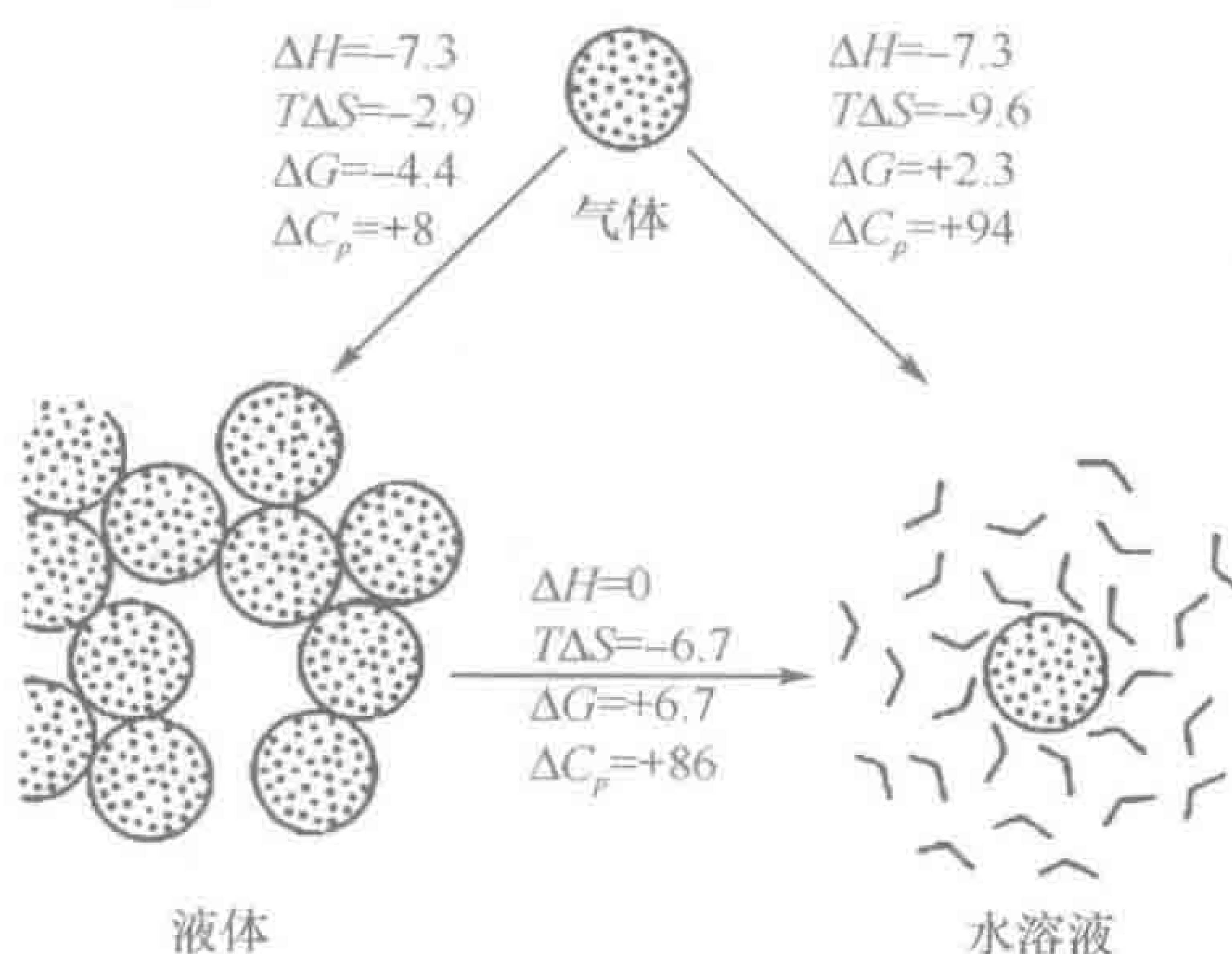


图 5-1　典型非极性分子环已烷 $(CH_2)_6$ 从气态(上)迁移到水中(下右)、迁移到液态(下左)以及从液态迁移到水中的热力学参数。数字大致对应于 25℃(即 $T\approx300$K，此时 $RT\approx0.6$kcal/mol)下的过程。ΔH(与能量一样，以 kcal/mol 为单位)为在所研究的迁移中每摩尔分子的焓的变化量；ΔS 为相应的熵变化量($T\Delta S$ 为熵变在自由能中的贡献；同样以 kcal/mol 为单位)；$\Delta G = \Delta H - T\Delta S$ (kcal/mol)为每摩尔分子在迁移中的吉布斯自由能变化；ΔC_p [cal/(mol · K)]为每摩尔分子在转变中的比热容变化。图片取自文献[6]，有修改

■ 于是,图5-1中所列的数字显示,就**能量**而言,$(CH_2)_6$分子与水的吸引和它与液态环己烷一样,都是能量为$\Delta H = -7.3$kcal/mol,但是它们不愿意进入水中(却很乐意进入液态环己烷中)。这种疏水的原因,正如图5-1中所显示的,是熵:$(CH_2)_6$分子浸入水中时,熵降低很多……

为什么非极性分子,如CH_4或$(CH_2)_6$从水中逃离?原因在于:H_2O分子能结合形成氢键,而非极性的H_2或Ar,CH_4或$(CH_2)_6$却不能。这也可以从众所周知的事实得到证实:乙醇CH_3CH_2OH,也能(像H_2O)结合形成氢键,它不从水中逃离……

可能有人会天真地提出,说原因在于CH_4或$(CH_2)_6$浸入水中会破坏H_2O分子的氢键。但根本不是这么简单!如果$(CH_2)_6$浸入水中只是破坏其中的氢键,那么在$(CH_2)_6$进入其中时溶液的能量会***剧烈升高***,但实际上它却**降低**(图5-1):每进入1mol $(CH_2)_6$为7kcal……不是能量升高,而是***熵减少***(强烈降低:$T\Delta S = -9.6$kcal/mol!)。

正是熵的减少阻止环己烷溶解于水中:一切都由自由能$G = H - TS$解决,它不仅随能量的增加而增加,而且随着S的**降低**而增加(S在G中带有负号!)。因此,甚至在能量H有所减小时,熵S的强烈减少都会导致自由能G的增加,结果是[式(5.1)]导致分子处于相应状态的***概率降低***(或者说,它们在该状态的密度降低)。

■ 现在我们就遇上了一个重要的物理问题:为什么水的熵会降低?

因为,对每一个与非极性分子的非极性表面紧密相贴的H_2O分子,非极性分子以其表面围住了由水填充的部分空间,而如果该水分子以自己的H原子指向该表面,则非极性分子有切断该分子的氢键的危险(图5-2)(提醒一下,氢键是有方向性的,它们只能在特定的分子朝向结合:当O—H键指向另一个水分子的O原子时才行)。

还请回想一下,在水中几乎所有可能的氢键都结合了,因此牺牲它们的结合可能性而不使水的自由能出现巨大损失是不可能的。而每一个这样的键都需要O—H基团直接指向与之成键的O原子。为了避免损失氢键,H_2O分子转变朝向(图5-2中上方的分子),并部分地在疏水表面冻结自己的热运动。它们并没有失去氢键,但为此付出了自己的熵。在这种情况下每摩尔近表面水分子自由能损失约为0.2kcal/mol,比可能的氢键损失(5kcal/mol)低一个数量级。

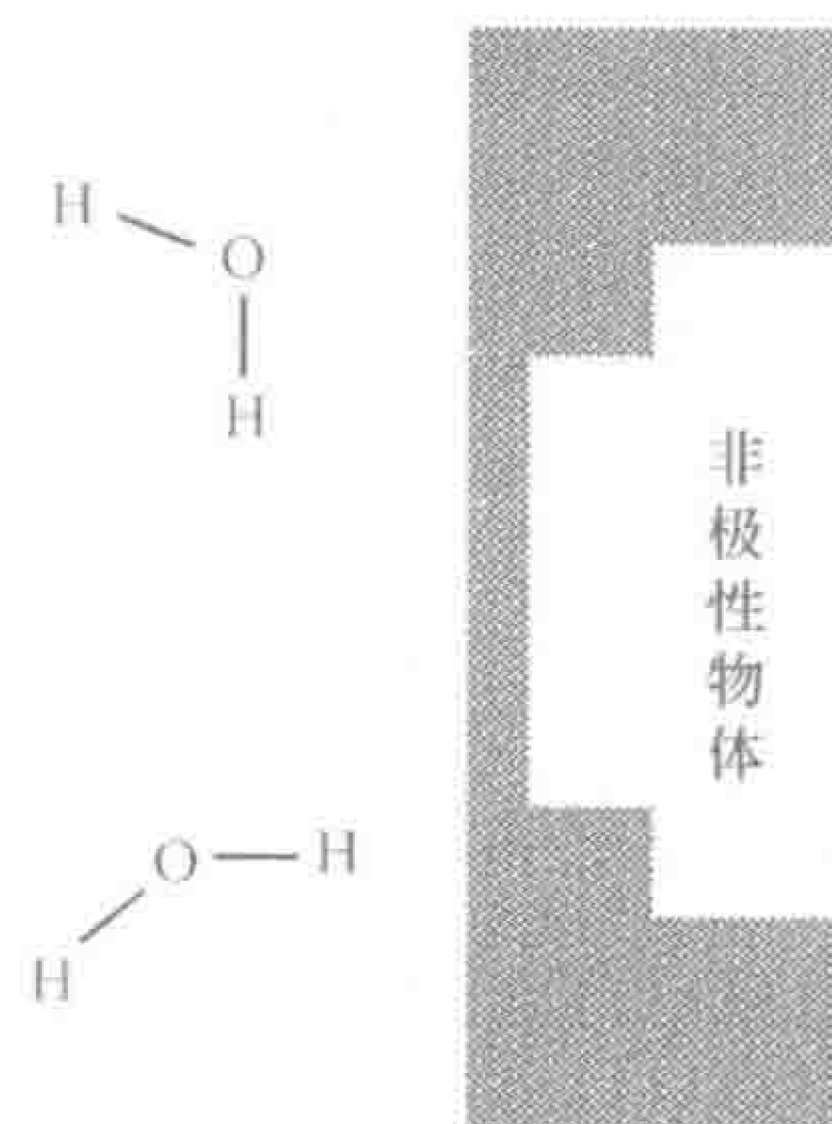


图5-2 非极性物体表面附近的H_2O分子。上方的水分子可以结合两个氢键,但由于其相对于物体表面的"贴近"分布导致熵减少。下方的水分子失去了一个与水分子形成的氢键:它的O—H基团指向非极性障碍物

在温度较低时,部分冻结在非极性物体表面的H_2O分子还能较容易地以氢键彼此连接。然而,如果温度很低(小于20℃,对许多碳氢化合物则可达60℃),近表层的氢键

甚至在能量上可以得到一点好处(现在它们较少被半冻结的近表层的水分子的运动所变形),但这种好处并不能补偿这些近表层分子的冻结而伴随的熵损失。

请注意:这种效应又一次具有熵性质,而不是能量性质,这恰恰**是因为**氢键的能量非常高:因为如此,H_2O 分子宁愿冻结(尽管这同样是热力学不利的)并损失部分自由度,即部分熵,而不愿失去氢键的较大能量。

我想强调一下,总的熵效应与天真地预期的能量效应具有**相同的符号**,但数值上要**小**。

在非极性物体表面的 H_2O 分子是被冻结的,也可以由环己烷(以及其他烃类)在水中的异常高热容得到证实。在水环境中 $(CH_2)_6$ 的热容(依赖于分子间相互作用)是在自己的兄弟(环己烷)中的 10 倍(!)(图 5-1)。更准确些,高热容并不是烃本身的,而是它在水中的与冰类似的水合外壳的:这个冰块在温度升高时会融化,这就使它具有异常的高热容。

在这里应该记住,以氢键结合并冻结的表层 H_2O 分子**并不完全**像普通冰中那样相互指向。在冰中分子应该规则地排列在晶格的各个顶点:它们必须形成巨大的三维晶体。而在表面它们可以随意排列,只要能结合形成氢键就行。这些分子不必遵守晶格的平移对称性,因为它们所形成的微晶粒、微冰块不必增大到无限大,它们只需以薄薄的一层包裹住浸入其中的疏水分子(图 5-3)或这些分子的基团就行了。在后一种情形下在这些分子之间就会形成**疏水键**。

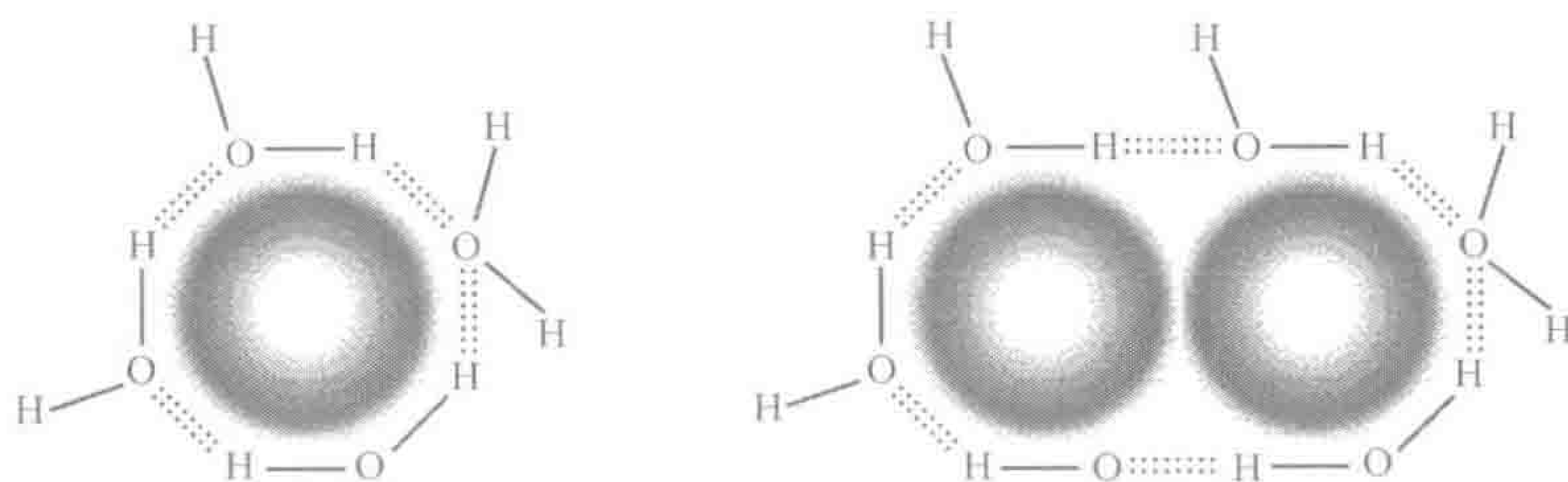


图 5-3 以氢键结合的水分子围绕非极性分子(左)和围绕一对同样的分子(右)的不规则包裹的示意图。在后一种情形下形成了疏水键

在疏水效应下水的极端有序化情形可用所谓的笼形包合物 (clathrate)(由水和非极性分子组成的晶体)加以说明。正如你们从化学课中知道的,笼形包合物远不如由水和极性分子组成的结晶水合物那么稳定。笼形包合物只能在低温(接近0℃)和高压下才会出现,这时(由于压强的作用)许多非极性气体分子被插入到水中。在笼形化合物中 H_2O 分子自身中的氢键达到饱和,就像冰中一样,但是这种冰具有与普通冰不一样的几何形状。在这种情况下会形成晶体,类冰把规则地分布的气体分子密封并维持在自己的核中。顺便说一句,有人认为,在笼形包合物中比通常的天然气田中包括更多的天然气,从笼形包合物矿床(实际上,它们很深,巨大的压强才能允许笼形化合物存在)中开采气体将可能是未来的工作。

■ 疏水效应的温度依赖性相当强(图 5-4)。温度并不是强烈地影响自由能 ΔG 本身,而是强烈地影响其分量 ΔH 和 ΔS 的值(甚至正负号!)。

在温度升高时表面层的氢键会熔化,但由此产生的疏水效应开始只是增加(高至 140℃),因为热力学不利的表面水分子有序化会停止,而热力学有利的氢键被破坏。

如果在室温下疏水效应仅仅是由熵产生的,则在温度升高时损失的(在非极性分子表面)氢键的能量会在这种效应上开始起着越来越大的作用。但疏水效应本身(由 ΔG 值决定)在 $T\Delta S < 0$ 时(即直到 140℃)总是会增加。

温度更高时,在水中(这时它只在有高压强下才能保持液态)有太多的氢键断裂,疏水表面越来越少地阻碍它们形成,疏水效应开始减弱。

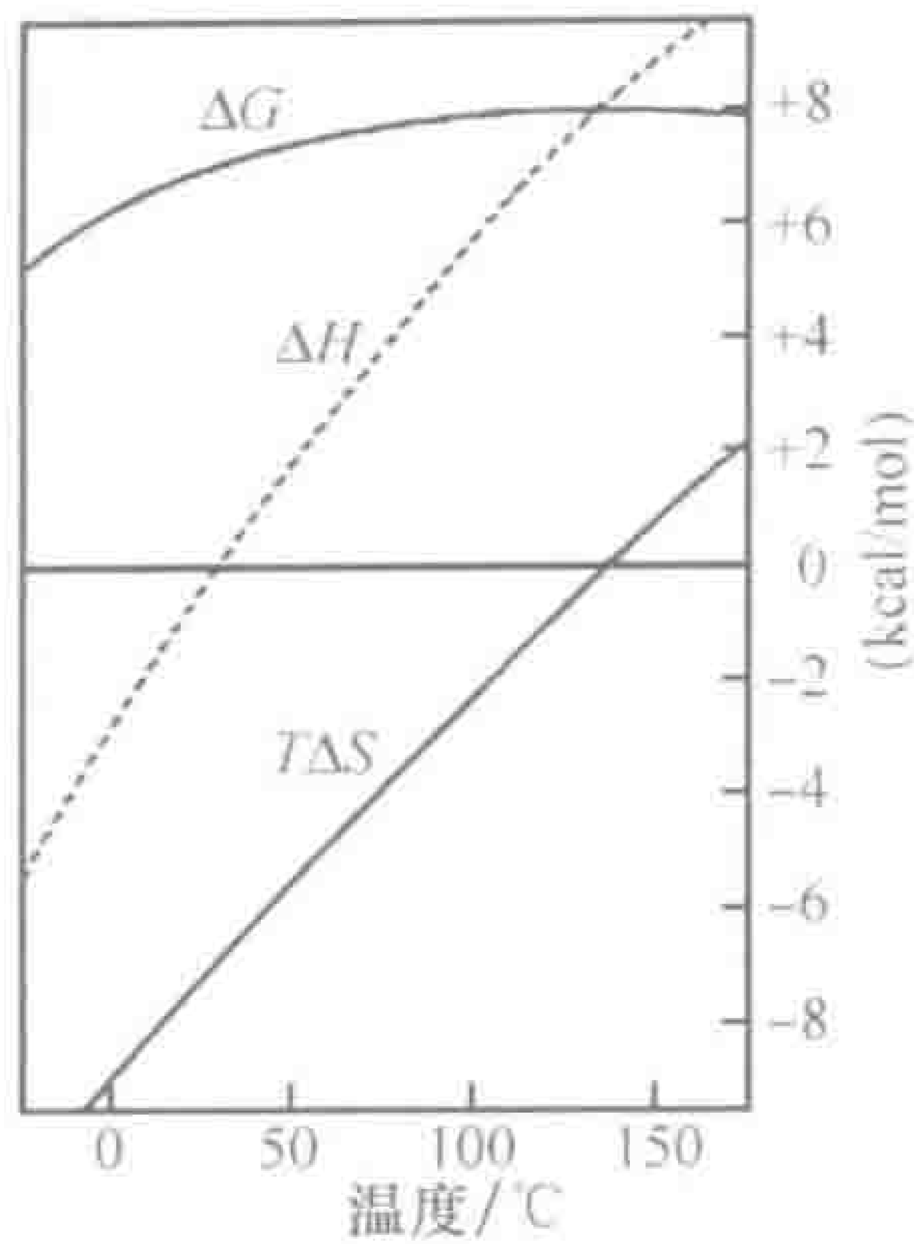


图 5-4 典型非极性分子戊烷 C_5H_{12} 在不同温度下由液态迁移到水中的热力学曲线。迁移自由能变化 $\Delta G = \Delta H - T\Delta S$,其焓变成分 ΔH 以及熵变成分 $T\Delta S$ 都以 kcal/mol 迁移分子为单位计量。ΔG 在 $\Delta S = 0$ 的点达到其最大值,而戊烷在水溶液中的饱和浓度(正比于 $e^{-\Delta G/RT}$)在 $\Delta H = 0$ 的点最小。图片取自 Privalov P. L., Gill S. J. 1988. Adv Protein Chem. 39:191-234,略有修改

在图 5-4 中我特意只列出了非极性分子从非极性溶剂(而不是从水蒸气中)迁移到水中的数据。因为我们感兴趣的是在蛋白质中的疏水效应,它们依赖于氨基酸残基从蛋白质核到水中的迁移,而非极性溶剂能比水蒸气更好地模拟蛋白质核。

■ 在实验上,氨基酸的疏水性是根据氨基酸在水和弱极性溶剂[常常是选用醇类(如辛醇)或二氧己环]之间的平衡分布常数而测得。这些实验相当复杂,因为一些氨基酸实际上不溶于水,而另一些却不溶于有机溶剂(如极性氨基酸根本不溶于纯环己烷或苯中)。不得不特意选择弱极性溶剂(能满足极化程度各不相同的氨基酸)以代替非极性溶剂,并使用其他技巧。因此根据所使用的弱极化溶剂,获得了多种数值发散的结果,尤其是对带电的或强极化氨基酸。不过定性上它们还是互相符合得不错。

关于氨基酸的疏水性我们将会稍迟一点讲述,现在来看一些数字是有益的,这些数字显示了与蛋白质中相似的非极性基团在室温下的典型疏水效应的特征。

从表 5-1 很快可以看出,ΔG(与 ΔH 和 ΔS 不同)随疏水分子的尺寸而增大。增大,但具体如何?对各种非极性分子的疏水性的更详细分析表明,它们的疏水自由能 ΔG 的增大大致正比于疏水分子的**水可及表面**。水可及表面的物理意义和构建方法可用图 5-5 说明。

表 5-1　25℃时疏水基团从非极性液体迁移到水溶液中的典型热力学参数

物质	迁移	ΔG /(kcal/mol)	ΔH /(kcal/mol)	$T\Delta S$ /(kcal/mol)	ΔC_P /[kcal/(mol·K)]
乙烷 $(CH_3)_2$(相似于 Ala 侧链基团 —CH_3)	苯→水	+3.6	−2.2	−5.8	+59
	CCl_4→水	+3.8	−1.8	−5.4	+59
苯C_6H_6(相似于 Phe 侧链基团—CH_2—C_6H_5)	苯→水	+4.6	+0.5	−4.1	+54
甲苯 $C_6H_5CH_3$(相似于 Phe 侧链基团)	甲苯→水	+5.4	+0.4	−5.8	+63

数据取自 Tanford C. 1980. The Hydrophobic Effect.

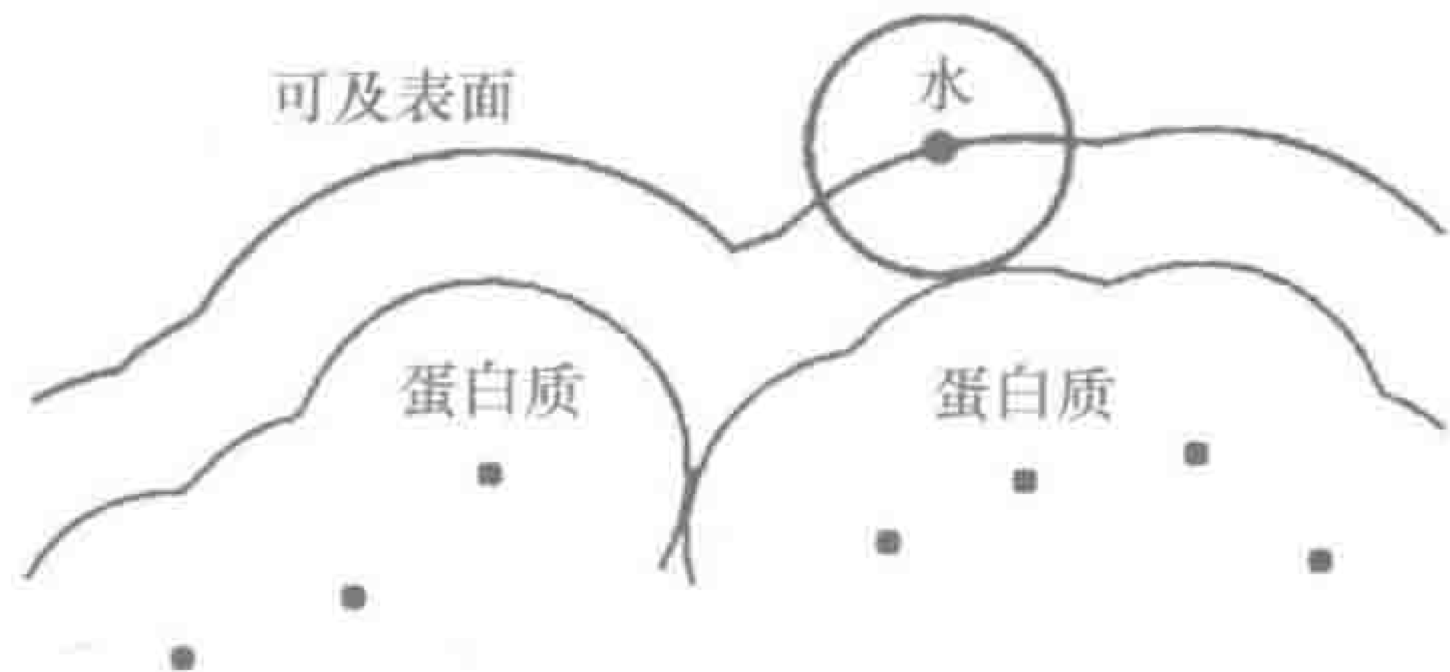


图 5-5　浸于水中的分子的水可及表面。点表示与水接触的原子的中心;连续线表示它们的范德华壳层。水分子以半径为 1.4Å 的小球表示。"水可及表面"是由这个小球在浸于水中的分子上沿其内部原子的范德华表面滚动而定义的。图片取自文献[3],有修改

非极性分子从非极性溶剂迁移到水中时,其疏水效应的自由能约为每平方埃米"可及表面"0.02~0.025kcal/mol。例如,对于那些我们讲过的分子有:苯的可及表面为 200Å^2,$\Delta G\approx 4.6$kcal/mol;环己烷的可及表面为 300Å^2,$\Delta G\approx 6.7$kcal/mol。

在疏水氨基酸残基上也观察到同样的规律性(图 5-6)。

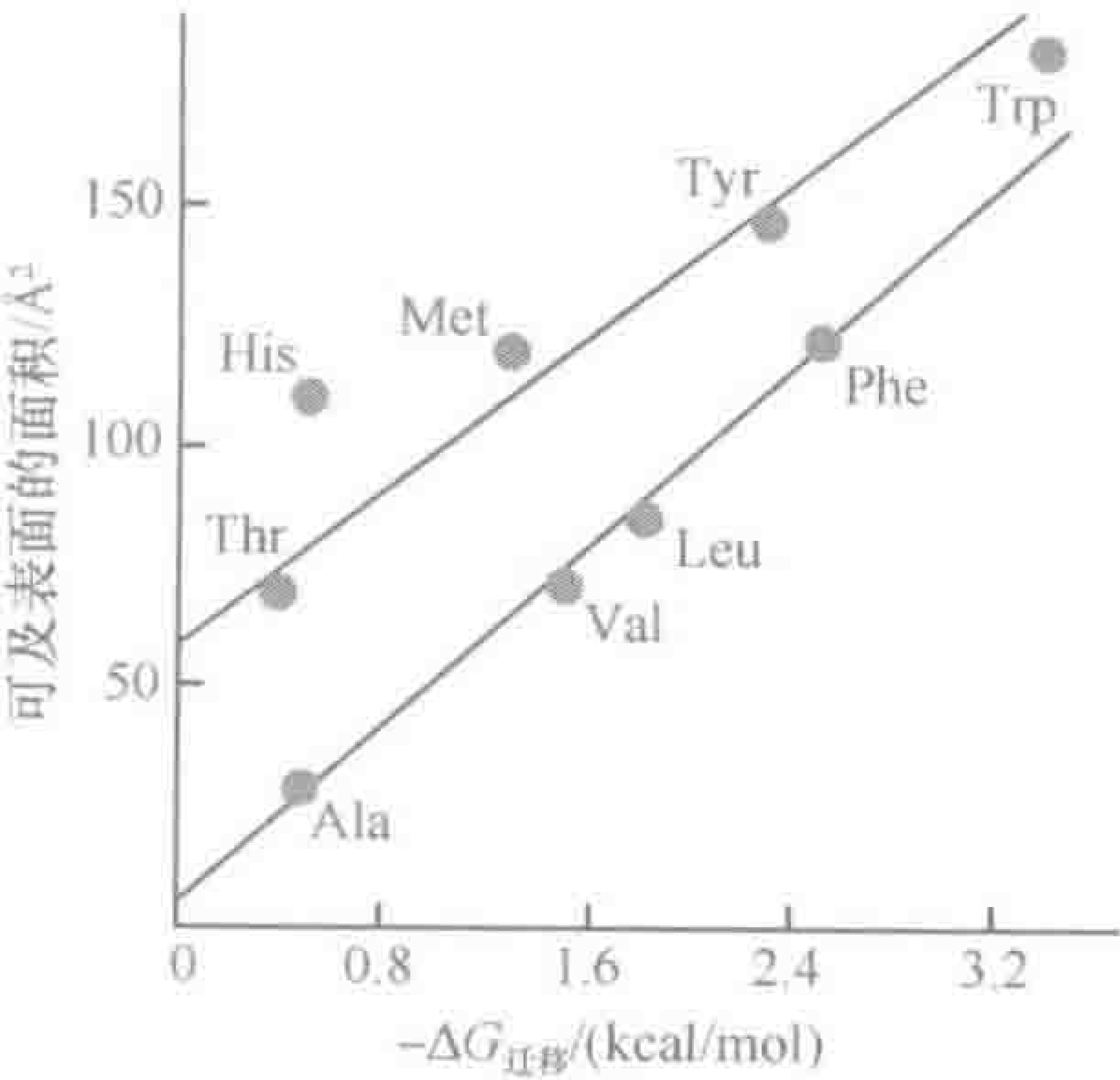


图 5-6　氨基酸残基的水可及表面面积和这些侧链的疏水性。(残基 X 的侧链可及表面等于氨基酸 X 的可及表面扣除无侧链的 Gly 残基的可及表面;残基 X 的侧链疏水性等于实验测得的残基 X 疏水性减去残基 Gly 的疏水性)。残基 Ala、Val、Leu 和 Phe 具有纯碳氢侧链。Thr 和 Tyr 在侧链上还有一个 O 原子,Met 有一个 S 原子,Trp 有一个 N 原子,His 有两个 N 原子;因此它们的**非极性**可及表面的面积小于整个可及表面的面积。图片取自文献[3],有修改

带有极性原子的侧链基团的疏水效应(在总的水可及表面相同时)比纯非极性基团的要小。然而,如果只考虑由非极性原子形成的水可及表面(即从总的水可及表面按每个极性原子减去大约 50Å^2),可得所有基团大致具有同样的侧链基团表面依赖性。

■ 疏水效应对维持蛋白质结构的稳定性起着重要的作用。正是它使蛋白质链折叠成为密实的球体。图 5-7 表明,疏水基团从水中迁移到疏水液体的自由能非常高:每摩尔几千卡,而该非极性液体在生理温度下的凝固自由能接近于零。[更准确点:对于相对小分子的碳氢化合物,例如,图 5-7 中显示的环已烷,凝固自由能甚至是正的,但这并不会使它在室温下变硬。凝固被熵所抵抗,尤其是分子在液体中的转动和平动的熵(在液体中每一个分子都或多或少地自由游动,而在固体中它们被晶体点阵所固定。粗略地说,分子的平动熵并不依赖于其大小,这与焓不同)焓随着一个分子与其他分子的连接数的增大而增大,即随着分子表面的增大而增大。在蛋白质链中运动熵的作用较小,因为氨基酸残基连接成链,即不能互不相关地运动,这有助于蛋白质的凝结。]

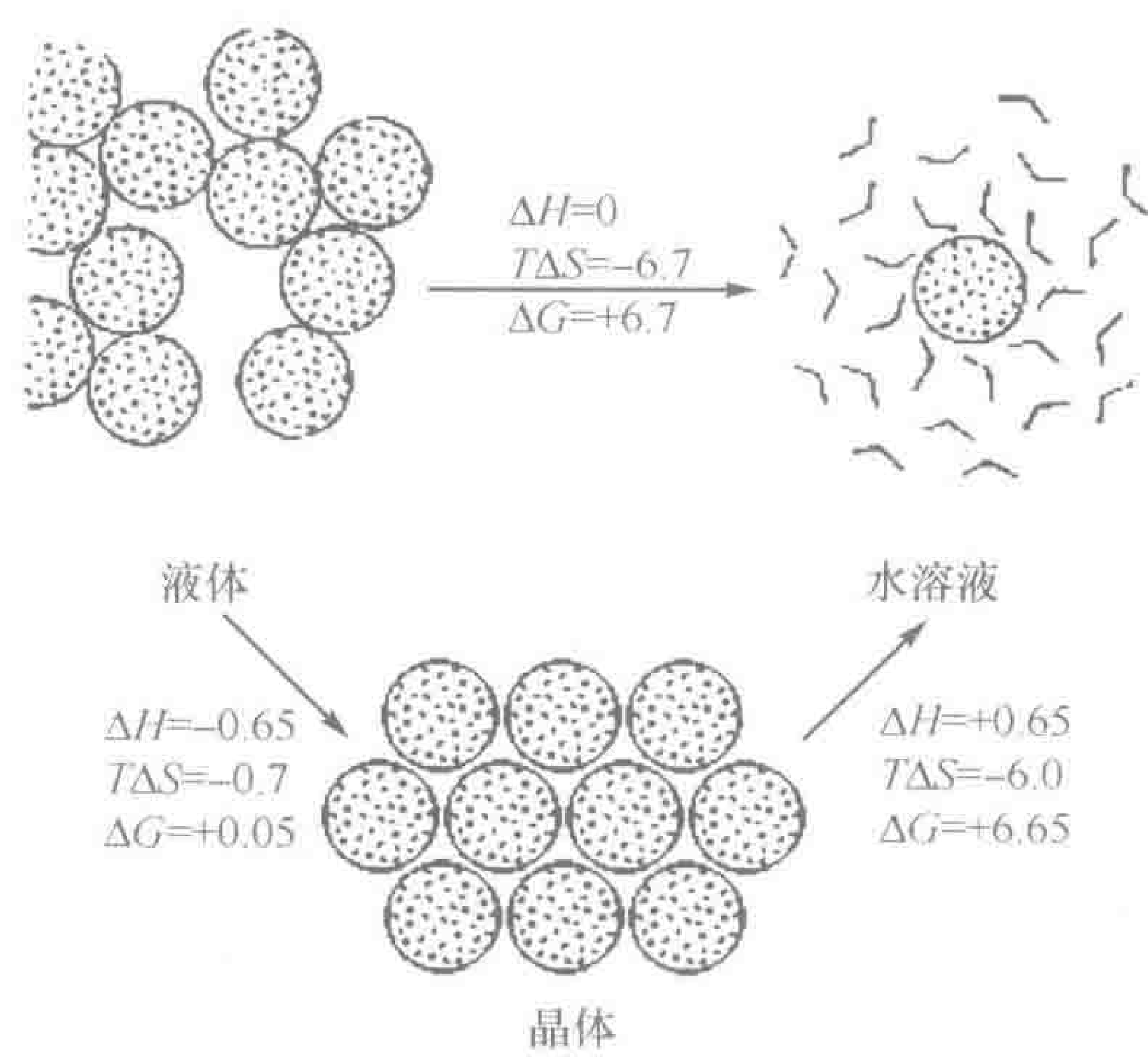


图 5-7 典型非极性分子环已烷从液相转变为固体和迁移到水溶液的热力学。数字对应于 25℃时的过程。ΔH、$T\Delta S$ 和 ΔG 值以 kcal/mol 给出。图片取自文献[6],但所有数字是根据参考书和式(5.10)~式(5.15)重新计算得更准确

甚至如果在环已烷的凝固自由能 $\Delta G_{液态环已烷-固体}$ 中忽略它的全部结晶熵,即假设 $\Delta G_{液态环已烷-固体} \approx \Delta H_{液态环已烷-固体} = -0.65$kcal/mol,则也看得出,环已烷的凝固热力学效应比把溶于水中的分子冷凝为液滴时所产生的疏水效应弱得多,因为 $\Delta G_{水-液态环已烷} = -6.7$kcal/mol。

■ 因此,粗略地说,疏水效应为蛋白质球体的形成贡献了 90% 的工作。但光靠疏水效应并不能形成天然固态蛋白质。它只是生成熔融的蛋白质球体,对此以后将要谈到。像所有有机液体一样,蛋白质的硬化是在范德华力及氢键和离子键的作用下发生的,它们是更专一的、比简单的疏水作用对原子结构的细节更敏感的相互作用。但那是对天然蛋白质的最后打磨,而大部分工作、所有的基本工作都归功于疏水效应。

第6讲(两倍)

水环境对静电相互作用的影响。蛋白质表面和内部的静电场。介电系数。盐溶液中的电荷屏蔽。利用蛋白质工程测量蛋白质中的电场。二硫键。配位键。

这一讲我们将要讨论静电相互作用,尤其是存在有蛋白质球体和其周围的水时给它增添的特殊性质。初看上去,这里所要讨论的是:你们都记得,在介电系数为 ε 的媒质中,电荷 q_1 在距离为 r 处所产生的电场的电势为

$$\varphi=\frac{q_1}{\varepsilon r} \tag{6.1}$$

而与处于该位置的电荷 q_2 的相互作用能量为

$$U=\varphi q_2=\frac{q_1 q_2}{\varepsilon r} \tag{6.2}$$

我想,你们也知道,在真空(或空气中)ε= 1,在水中 ε 接近于 80,而在塑料类的媒质中 ε 介于 2 ~ 4。

课堂里的声音:这都正确,但只是"大体上"正确。首先,当电荷不是处于真空中,而是处于某种媒质中时,则严格地讲,U 不是能量,而是**自由能**[因为,如果我们的系统(电荷)与媒质有热交换,在系统趋于平衡时,正是自由能趋于最小值,即应该是自由能负责在异种电荷间的吸引和同种电荷间的排斥]。此外,基本相互作用的能量不应该依赖于温度,而 U 依赖于介电系数 ε,因而也会随温度而变化。

讲课人:这个意见完全正确。但所有这些都是琐碎细节,是吹毛求疵,因为我们感兴趣的是电荷相互作用对蛋白质稳定性的影响。

更重要的是,式(6.2)只有在同种均匀介质中才是正确的。当我们对蛋白质中的电荷相互作用感兴趣时,我们是在处理极不均匀的介质。蛋白质本身,与所有塑料一样,其 ε 值并不大,大约是 2 或 4(提醒一下:在空气中 ε=1)。而在水中,ε=80。蛋白质的带电基团通常分布在它的表面,就在水的旁边(后面我们会看为什么会这样)。当我们估算蛋白质中的静电相互作用时,ε 的值应该选取为多大呢?如果我们取 ε≈80,两个单位电荷(质子)在相距 3Å 时的相互作用能量 ≈1.5kcal/mol;如果取 ε≈3,则它们的相互作用能量 ≈40kcal/mol。差别非常大:多出来的 +40kcal/mol 会破坏任何蛋白质结构……

顺便说一下,如何估算,对蛋白质是否存在这种或那种(自由的)能量的作用?这里有必要记住两个数值:①kT≈0.6kcal/mol;低于 kT 这个值(每个蛋白质分子

上）的任何作用永远都不会存在，它们会被热涨落所“抹去”；②典型的蛋白质结构“稳定性安全系数”（等于蛋白质天然态与变性态之间的自由能）≈10kcal/mol；超过 10kcal/mol 的作用会导致蛋白质“爆裂”。

第二个问题也与此相关。当电荷间的距离 r 远远大于媒质分子尺寸时，式（6.1）和式（6.2）是等价的。但是在蛋白质中电荷通常处于直接接触中，相距 3～4Å，这时甚至水分子，更不用说侧链基团，都无法插入到它们中间。在这种情况下该如何估算静电相互作用？像在真空中那样取 $\varepsilon=1$ 怎么样？或者都取 $\varepsilon=80$？或者……？

几句哲学性的题外话。为什么一般总是需要这种大致的估算？粗看上去，如果拥有功能强大的计算机，可以向里面输入“所有东西”：液态水分子、蛋白质原子的坐标、电荷坐标，设定温度（即热运动的能量），就“全部都能准确计算出来”。实际上这是一个空想的图景。计算，我指的是细节的计算（由所谓的“分子动力学”来进行），要进行几小时或者好多天（顺便说一下，它不可能完全精确，请回忆一下，由于 p 轨道的缘故，原子不是圆的）：因为你需要考虑热运动和数十万个原子的极化。而你们，确切地说，感兴趣的是简单、但是快速的估算：是否可以在蛋白质中某个位置引入一个电荷，这时蛋白质是否会爆裂？我的目标，就是教给你们进行这种估算。

■ 首先，来估算一下，当电荷从水中（$\varepsilon\approx80$）移动到蛋白质中（$\varepsilon\approx3$）时，它的能量如何变化。微粒性，即蛋白质和水的原子结构，我们暂时忽略，更准确地说，是把这个问题暂时放到一边。

根据经典静电学，半径为 R、带电荷 q 的小球在介电系数为 ε 的媒质中具有能量

$$U=\frac{q_1^2}{2\varepsilon R} \qquad (6.3)$$

这个公式直接由式（6.2）推得：当我们给小球充电（由带电量为 0 充到带电荷 q_1），每移动微小的电荷 dq 到它的表面，根据式（6.2），每个元电荷 dq 使小球的能量增加 $dU=q\,dq/\varepsilon R$，积分 $q\,dq$ 从 0 到 q_1 得到 $q_1{}^2/2$。



$\Delta U_1=2\Rightarrow q^2/2\varepsilon_2R-q^2/2\varepsilon_1R$

带电原子的径约为 1.5Å。这意味着，它的（自由）能在 $\varepsilon\approx80$（在水中）时约为 1.5kcal/mol，在 $\varepsilon\approx3$（在蛋白质中）时约为 40kcal/mol。这个很大的差值说明了，为什么在蛋白质内部（不同于其表面）实际上没有带电基团（不难估算，把紧密靠近的一对异号电荷浸没到 $\varepsilon\approx3$ 的媒质中增加的自由能同样是大约 40kcal/mol）。因此，浸入在蛋白质深处的带电侧链基团总是被放电的，即把自身的多余 H^+（如果它原先带正电）送给水或从水中抓取到自身缺少的 H^+（如果原先带负电）。确实，带电基团的放电也会使自由能升高，但“仅仅”只有几（而非几十）kcal/mol。如果更严格地讲，是

有一些电荷存在于蛋白质内部,但是实际上在这种情况下它们总是有重要的功能性作用:蛋白质容忍电荷处在它的内部,是为了工作……

■ 现在来学习估算在考虑蛋白质($\varepsilon \approx 3$)与水($\varepsilon \approx 80$)的交界面时的电荷相互作用。

来看下面这道简单的教学习题。为了解题简便起见,让水($\varepsilon = 80$ 的媒质)占据空间的一半,"蛋白质"($\varepsilon = 3$ 的媒质)占据另一半,两者之间的交界面为平面。设在"蛋白质"表面上方的水中①处有电荷 q(图6-1)。求由这个电荷在任意点2处产生的电场 φ。

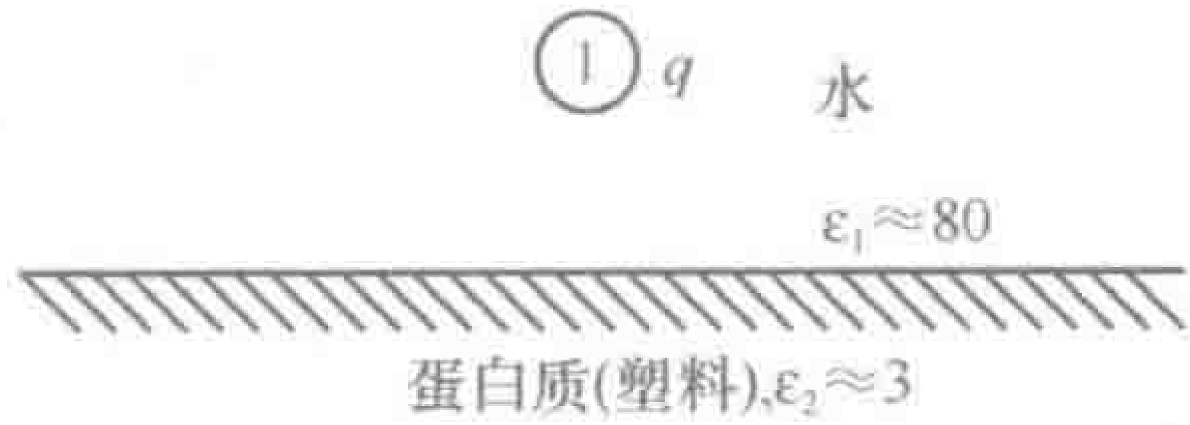


图6-1 蛋白质中的侧链基团的典型情形。位于靠近蛋白质球体表面的点①的电荷 q 被水所包围,即它自身处于高介电系数 ε_1 的环境中,但附近是低介电系数 ε_2 的环境

如果蛋白质不存在,或者它距离很远,则在任意点2的电场按照经典公式(6.1)进行计算,取 $\varepsilon = 80$(水的介电系数)和 $r = r_{12}$(点1和点2之间的距离)。

如果电荷距离蛋白质(介电系数不同的另一种媒质)很近,结果会如何变化呢?

作为提示,请回忆一下非常相似、你们可能在普通物理课程中求解过的问题。这个问题与前面基本一样,只是"第2种媒质"为金属,它是电流导体,即它的 ε_2 等于无穷大(图6-2)。这个问题的解如下所示:

$$\varphi = \begin{cases} 0 & \text{金属内部(无电场)} \\ \dfrac{q}{\varepsilon_1 r_{12}} + \dfrac{-q}{\varepsilon_1 r_{02}} & \text{金属上方} \end{cases} \tag{6.4}$$

金属**内部**电场为零,因为,如果不是这样,则它内部的自由电荷会移动,直到电场为零。

在金属表面**上方**(点2处)的总电场是由电荷 q **和**它在金属中的"**镜像**"所产生的。这个镜像⓪综合了在电荷①作用下金属中的位移电子所诱导的电荷的影响。这些电子分布在金属表面,产生了"镜像效果"。该"镜像"带有电荷 $-q$,位于点0处,该点在金属表面**下方**的深度与点1在金属表面**上方**的高度相等(因此,r_{02} 为镜像电荷与电场待求点之间的距离)。

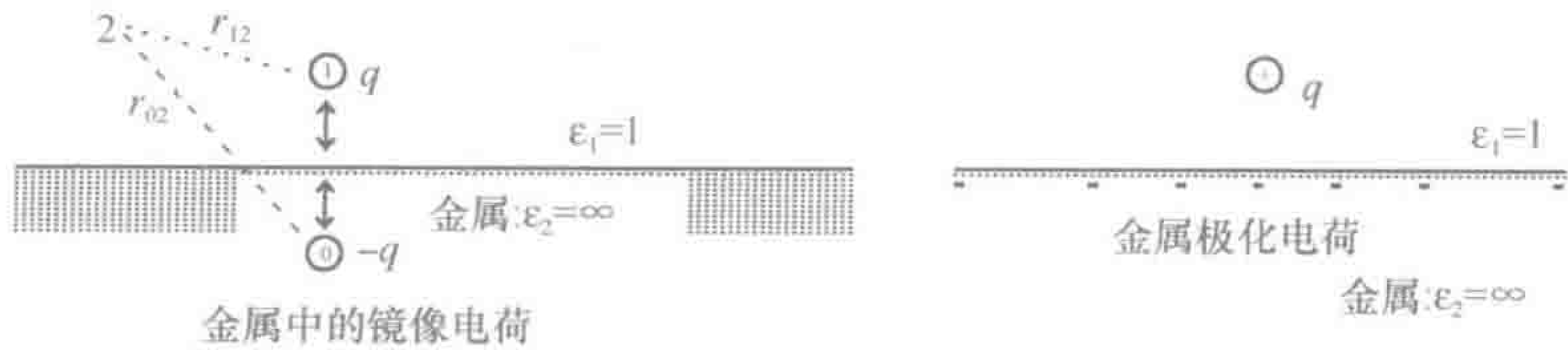


图6-2 关于电荷电场的中学补充习题的图解。电荷位置靠近高介电系数的环境("金属")。右图显示了"金属"中产生的极化电荷,左图显示了"镜像"电荷 $-q$,它综合了这些极化电荷在金属范围内的作用。详细解释见正文

我不想严格地证明这个结果（更准确说，是“金属上方的电场”的第二部分的结果，它并不是很明显），只是简要地叙述一下论证的基本思想。

在金属内部，如果既没有电流，也没有电场，则这个场的电场线（回忆一下，它们表示了作用在电荷上的力的方向，见图 6-3 中的箭头）应该完全垂直地进入金属（否则它们会有平行于表面的分量，会驱使电流沿金属流动）。但是由两个符号相反、电量相等的电荷（在不存在任何金属时）所产生的电场的电力线也会严格地垂直通过这两个电荷中间的平面。一旦电场的电力线外观相同，则产生的电势也完全相同。

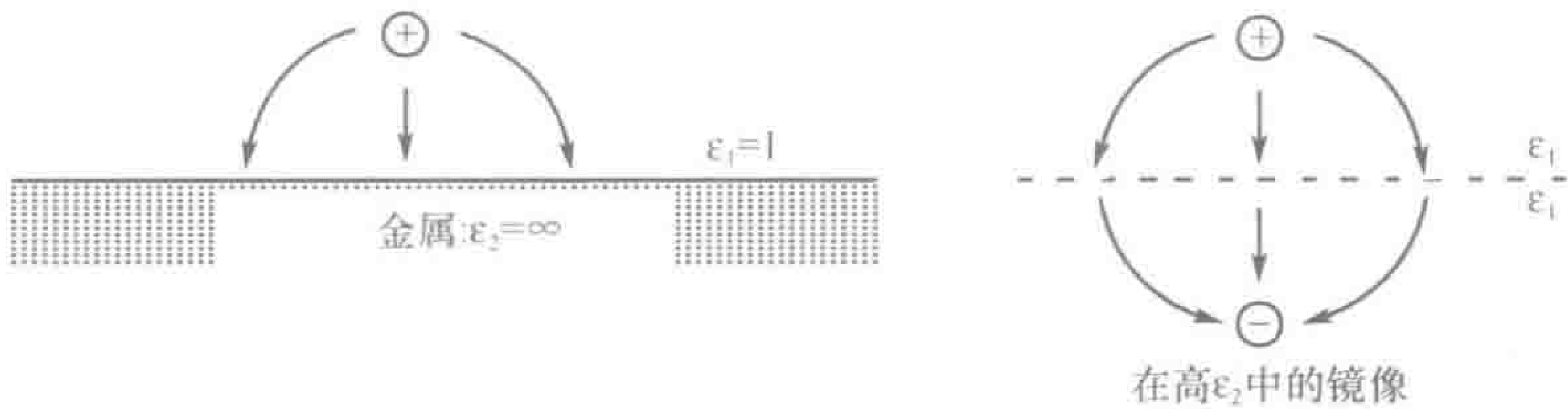


图 6-3　处于金属上方的电荷的电力线，以及数值相同、符号相反的电荷所产生的电场的电力线的相似（与上方部分相似）图景

“金属上方的电荷的电场”习题的解答提示，这类问题可以借助于“镜像”电荷和分析场中电力线的变化而方便地求解。而且这些线都趋向于**尽快地进入高介电系数的介质**。

现在回到我们关于处于两种介质交界面处的电荷所产生的电场的问题。在这种情形下高介电系数的介质——水——起着主要的作用［图 6-4(a)和(b)］。与水的大量极化电荷相比，蛋白质上少量的极化电荷（在蛋白质-水的交界面上）可以忽略不计。介质的极化导致电力线尽力留在水中（图 6-4，左下），因为它的介电系数比蛋白质要高。电场的这种形状（如果只看它位于分界面**上方**的部分）是由两个对应的电荷产生的［图 6-4(d)］。这提示，我们需要怎样的镜像电荷并把它放置在哪里：应该把它放在分界面下方，即我们观察到镜像的位置——而它的符号应该与我们（放置在蛋白质上方）的电荷“1”相同。

我不想拿所有习题的详细求解来把你们搞得很累，而是直接写出答案如下。

蛋白质上方的电场是由电荷 q 和镜像电荷 $q' = q\dfrac{\varepsilon_1 - \varepsilon_2}{\varepsilon_1 + \varepsilon_2} \approx +q$（在 $\varepsilon_1 \gg \varepsilon_2$ 时数值上接近于$+q$）所产生的。q' 处于交界面下方，深度等于 q 处于上方的高度（图 6-5）。

结果是，等效介电系数 ε_{eff}（即代入公式 $\varphi = \dfrac{q}{\varepsilon_{eff} r_{12}}$ 中的介电系数）在蛋白质上方区域只有与电荷①距离极近时（这时 r_{12} 比电荷①到蛋白质表面的距离小得多，$r_{12} \ll r_{02} \approx r_{01}$）接近于 ε_1，即大约为 80。

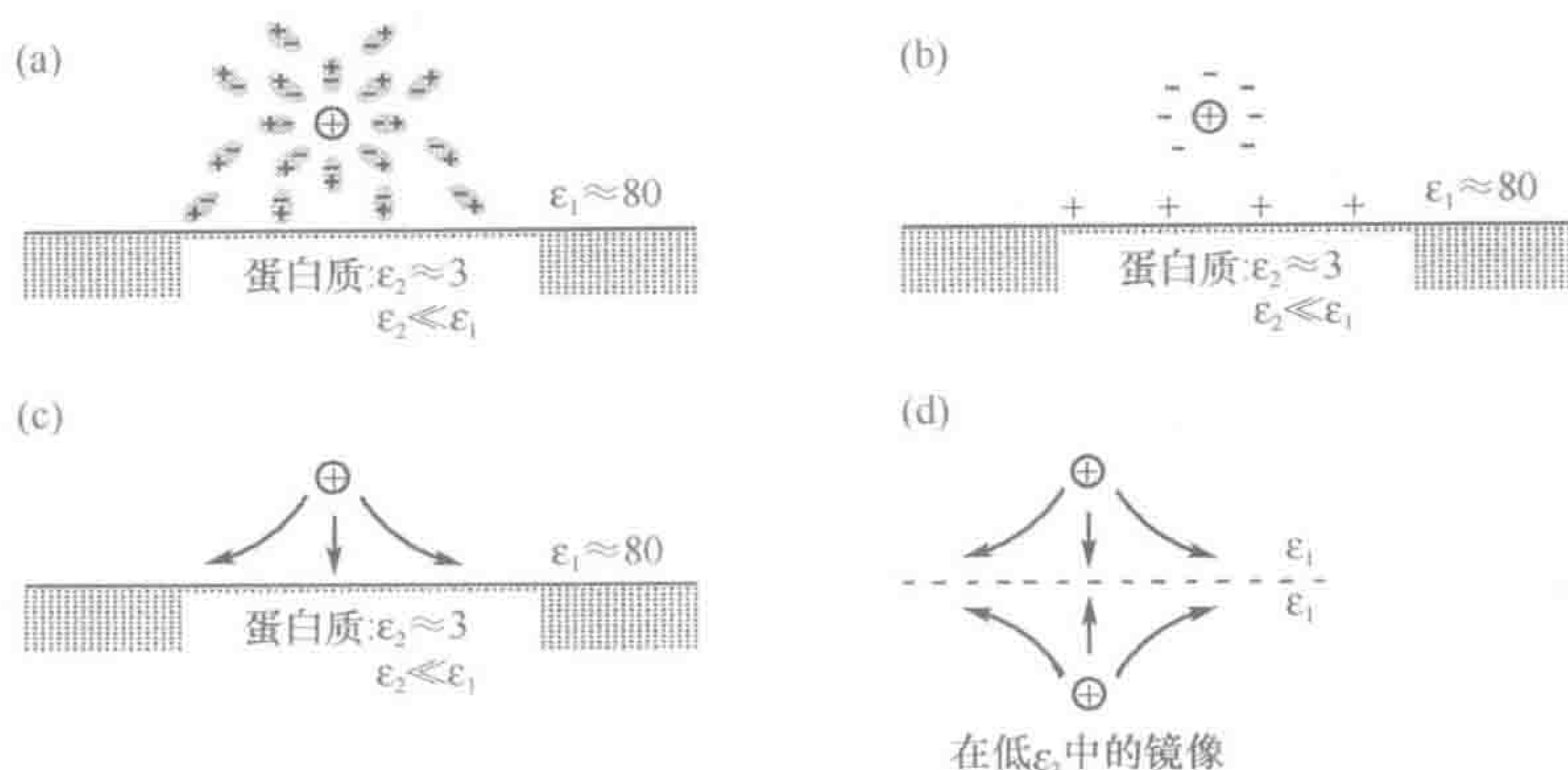


图 6-4　处于蛋白质表面附近水中的电荷的电场的习题(图 6-1)解答图解。(a)偶极子(水分子)把自己的“负号”转向我们的⊕电荷,而“正号”转向相反方向。(b)结果是在水中产生了极化电荷。因为内部水层中的水分子偶极子的“正号”和“负号”互相抵消,所以极化电荷出现在水的边界上。这些直接处于⊕电荷附近的极化电荷减弱了⊕电荷的电场,导致出现水的高电导率(这种效应与蛋白质的邻近无关)。位于分界面表面附近的水分子在该表面产生极化电荷:注意它与我们的⊕电荷符号相同。(c)在低(与⊕电荷所处的介质相比)介电系数介质的表面的电场电力线合成图。(d)由同符号的映像电荷产生的等价(从介质分界面上方的电力线走向的角度来说)图

但是在距离电荷①很远时 ε_{eff}总是大约为 40。实际上,当 $r_{12} \gg r_{10}$时,$r_{02} \approx r_{12}$,而且由于在 $\varepsilon_1 \gg \varepsilon_2$ 时有 $q' \approx q$,因此 $\varphi \approx 2q/\varepsilon_1 r_{12} = \dfrac{q}{\dfrac{\varepsilon}{2} r_{12}}$。

因此,蛋白质表面正上方(同样的,正下方)的场中(在该处 $r_{02} \approx r_{12}$,如图 6-5 所示),有 $\varepsilon_{eff} \approx 40$。

同样,而**蛋白质内部的电场**是由处在点“1”处的电荷所产生的(图 6-6)。对于该电场(我再一次省略计算过程),在交界面下方所有空间中的等效介电系数 $\varepsilon_{eff} = (\varepsilon_1 + \varepsilon_2)/2$,即大约为 40(即等于在蛋白质表面上方的 ε_{eff})。这个相等性是很自然的,因为“蛋白质内部电场”和“蛋白质外部电场”的电势在两种介质(水和蛋白质)的交界面上应该相等。

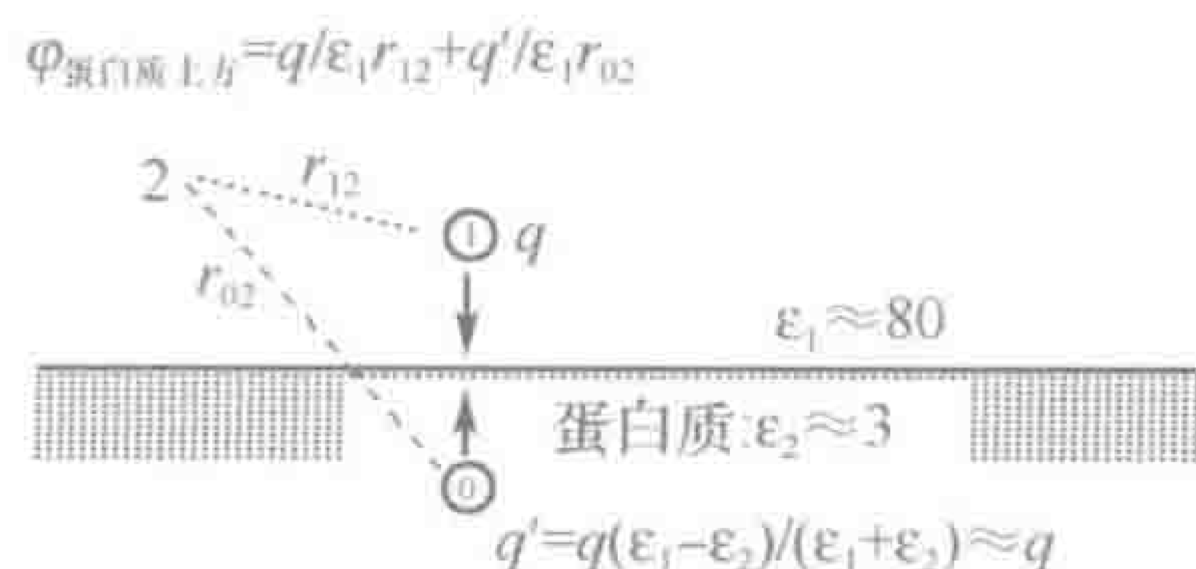


图 6-5　利用在蛋白质内的镜像电荷,计算蛋白质上方的电场

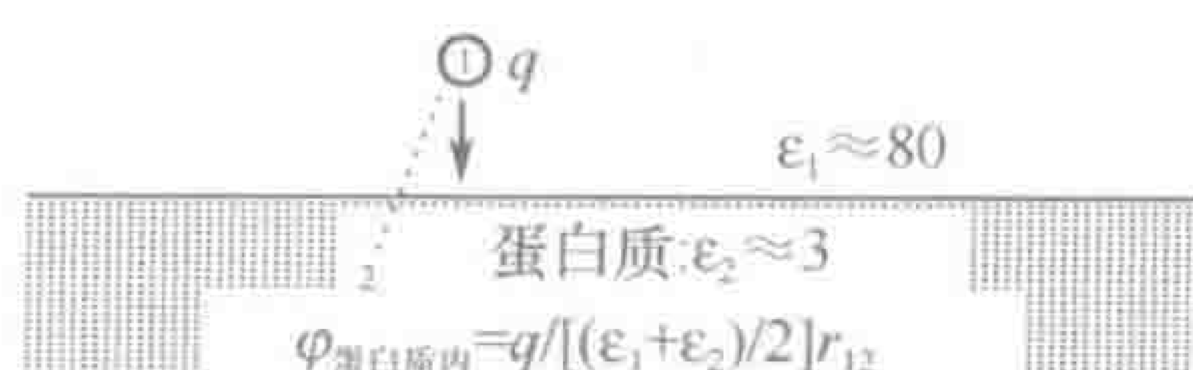


图 6-6　位于蛋白质表面上方的电荷在蛋白质内部产生的电场的计算

如果我们来看一个互补性的问题：计算处于蛋白质表面**下方**的电荷所产生的电场(图 6-7)，可以得到非常相似的结果。

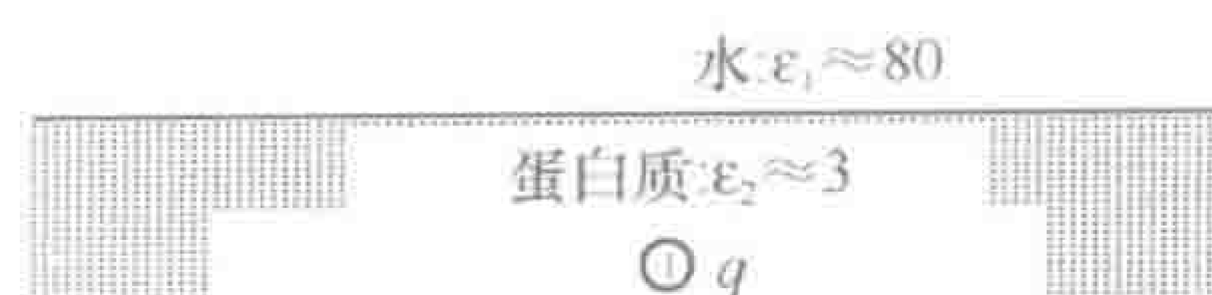


图 6-7　位于蛋白质内部，但靠近与水接触边界表面处的电荷的电场问题图解

在这种情形下在距离电荷 q 较远处的**任何**地方，等效介电系数 ε_{eff} 也都接近于 $(\varepsilon_1+\varepsilon_2)/2$，即大约为 40，而只有当 r_{12} 远小于电荷 q 到蛋白质表面的距离时，接近于 ε_2(即大约为 3)。

■　更有趣的结果来自这样一个问题：当电荷处于蛋白质一边时，它在蛋白质另一边所产生的电场如何(图 6-8)？

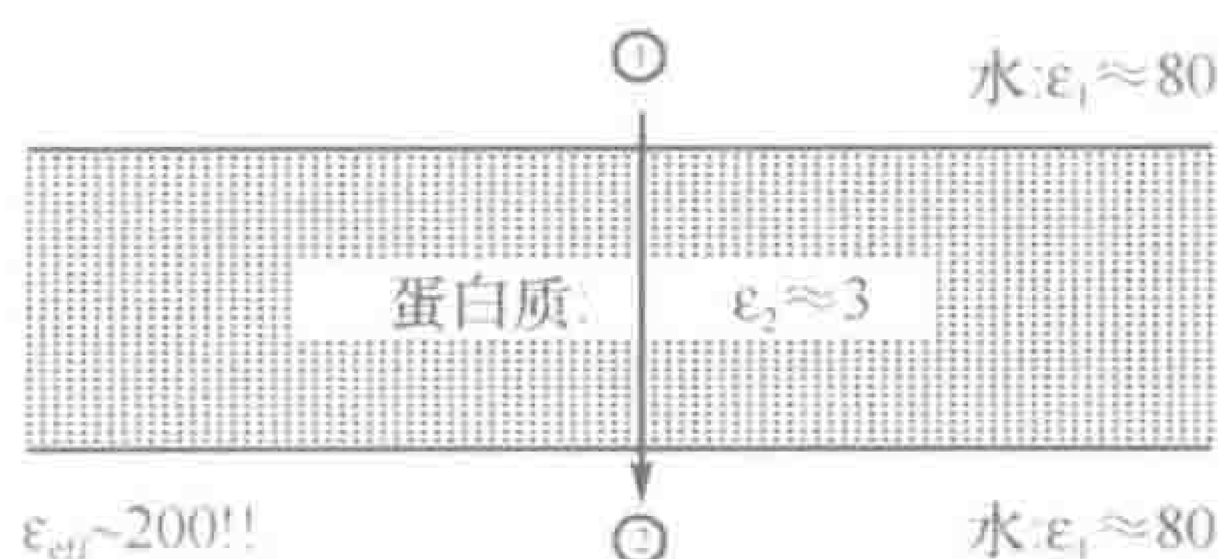


图 6-8　电荷①在蛋白质对面的点②处所产生的电场问题图解

看上去，在这种情形下，因为相互作用是穿过蛋白质这种低介电系数($\varepsilon_2\approx3$)的，我们应该可以期待，ε_{eff} 会接近于 3，在极端情况下(因为周围是水)，ε_{eff} 可能介于 3～80。但是实际上，ε_{eff} 达到了 200！为什么会这样？

为了解释这个问题，请看在这种情形下围绕蛋白质和电荷[极化水分子是如何指向的，以及这是如何改变电荷的电场的(如图 6-9 所示。为了清楚起见，图中的比例与图 6-8 画得不同，但是描绘的是同一现象]。

水分子的指向方向是沿着场的方向：它们的"-"主要是朝向我们的⊕电荷，而"+"朝向相反的方向。结果是，首先，⊕电荷部分地被相邻的水分子的"负电荷"所补偿；这是并不新奇的现象，它只不过是导致出现水的高介电系数 ε_1。第二，出现了我们感兴趣的现象。紧邻蛋白质朝向⊕电荷这一面的水分子的"正电荷"指向蛋白质，在这一面产生了极化"+"电荷；而紧邻蛋白质另一面的水分子的"负电荷"，产生了相反的极化"-"电荷(蛋白质本身所产生的极化电荷可以忽略不计，与水的极化电荷相比很小，因为它的 $\varepsilon_2\ll\varepsilon_1$)。

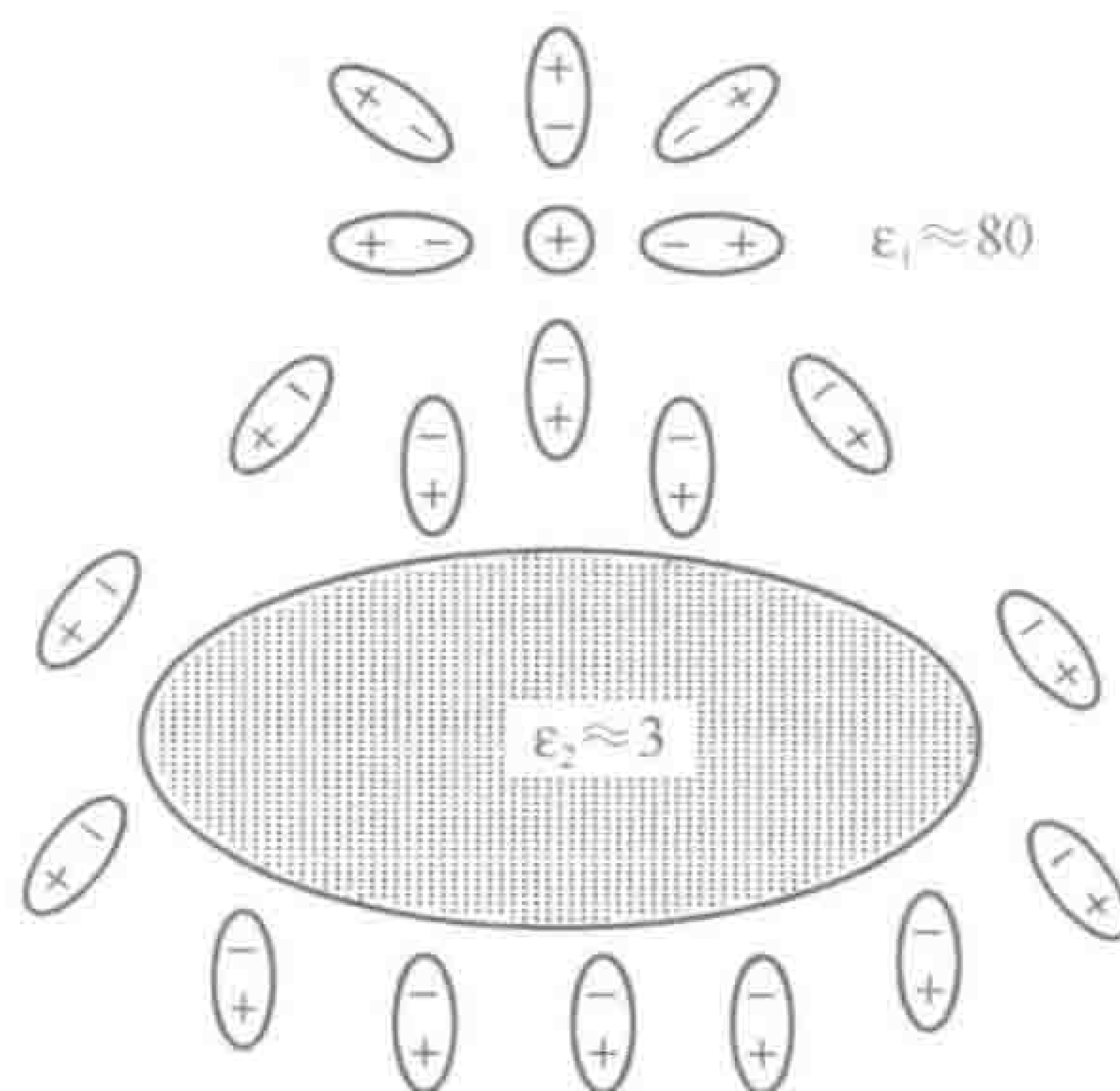


图 6-9　围绕蛋白质和⊕电荷(图示为正电荷只是为清楚起见)的水分子(图示为偶极子)的指向

结果就是产生了极化电荷：产生在"电荷-水"交界面和"蛋白质-水"交界面上。仅仅是在交界面上！在单一均匀的(初始不带电的)介质中极化电荷"中和为零"，在静电学中有这样一个定理。根据同样的定理，在被同一种介质(我们这里是水)所包围的物体(我们的蛋白质)的表面附近的外部电荷所产生的极化电荷总和也等于零。

这意味着，在靠近蛋白质表面的主要由转向水分子产生的总极化“+”电荷(！)指向⊕电荷，而与之相等的极化“-”电荷(由蛋白质另一面表面的水分子所产生)指向相反的方向。

结果是，在蛋白质朝向电荷的那一面，与不存在蛋白质时相比，电场的电势升高了。在该处，在⊕电荷的电势之外要加上在蛋白质这一面的表面的极化“正电荷”所产生的电势(而极化“负电荷”离这里较远，影响很小)。因此这里 $\varepsilon_{eff} \approx 40$，正如我们前面讲的一样。

同时，蛋白质背对电荷的那一面的电势与不存在蛋白质时相比，是**降低**了。这里在⊕电荷的电势之上加上了在靠近蛋白质表面的极化“负电荷”的符号相反的电势(极化“正电荷”距离此处较远，影响很小)。一旦这里的电势比不存在蛋白质时要**降低**，则这里(蛋白质背对⊕电荷的那一面)的 ε_{eff} 比不存在蛋白质时的 ε_1(即 $\varepsilon_1=80$)要大。

大体上，如果电荷①位于蛋白质表面附近，则由电荷①和它所引起的极化电荷所产生的电场在“蛋白质内部及其周围”的 ε_{eff} 值的分布，如图 6-10 所示。如果电荷①位于蛋白质内部，则如图 6-11 所示。

提醒一下，ε_{eff} 是用于代入计算电荷①在 r 点处的电势的公式 $\varphi(r)=\dfrac{q_1}{\varepsilon_{eff}\,|\,r-r_1\,|}$(其中 r_1 为电荷所处的点)中的 r 点处的等效介电系数。

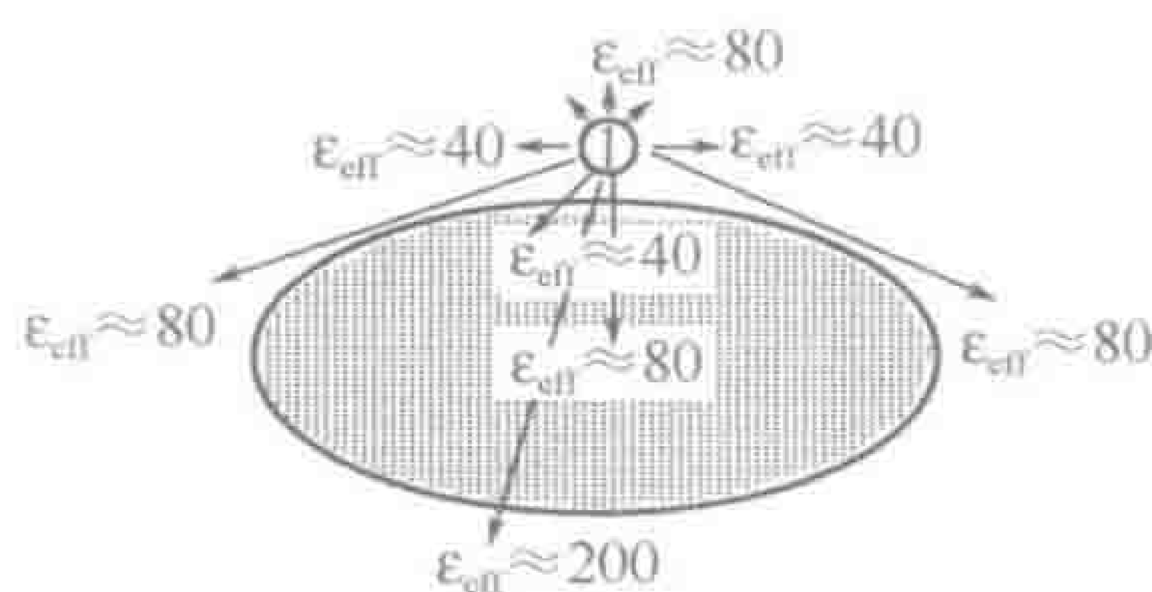


图 6-10　分布在蛋白质表面的电荷在不同 r 点处所产生的电势的 ε_{eff} 值的典型数值

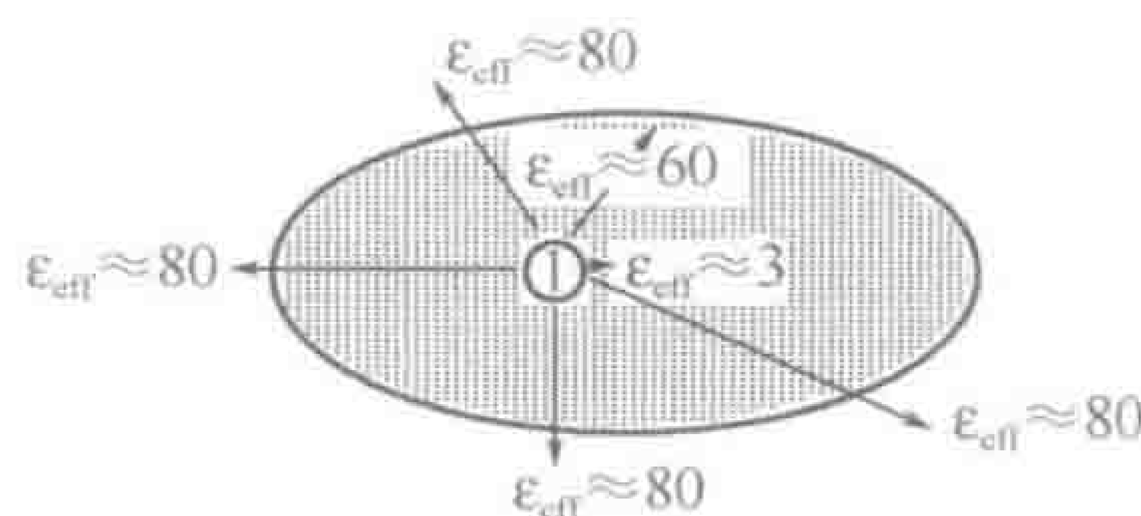


图 6-11　分布在蛋白质内部的电荷在不同 r 点处所产生的电势的 ε_{eff} 值的典型数值

还请注意一个由交界面引起的并与其“镜像电荷”相关的现象。它涉及电荷对自身的作用[也可参见对式(6.3)的讨论]：处于蛋白质外部的电荷被从蛋白质表面向外**排斥**，而处于蛋白质内部的电荷被强烈地吸引向其表面，即在这两种情形下，高介电系数的介质吸引电荷，而低介电系数的介质排斥电荷。这种吸引力和排斥力很容易估算：因为这里就是电荷与其镜像之间的相互作用(图 6-5，图 6-6)，我建议你们自己练习计算一下。

一个数值例子。可以证明，半径为 R、带电荷 q、一半浸于水中($\varepsilon_1=80$)而另一半浸于蛋白质中($\varepsilon_2=3$)的传导小球(离子)的能量是 $q^2/(\varepsilon_1+\varepsilon_2)R \approx 3\text{kcal/mol}$；全部浸在水中的能量为 $q^2/2\varepsilon_1 R \approx 1.5\text{kcal/mol}$[在蛋白质中是 $q^2/2\varepsilon_2 R \approx 40$，见公式(6.3)]。因此，位于蛋白质上的离子丢失了(与水中的离子相比)1.5kcal/mol。但是如果离子被夹在两个蛋白质中间，它总共减小的不是

3kcal/mol（如果它与这两个蛋白质的相互作用是累加起来），而是高出一个数量级：37kcal/mol！这表明，在非单一介质中的静电作用是**强烈地非累加的**。

■ 现在来看与微粒性，即介质的原子结构，相关的效应。

本质上来说，介电系数 ε 本身的数值是由介质的原子结构决定的。如果介质是由非极性分子构成的，电场只能移动其中的质子①，而这是很困难的：因此质子移动很小，ε 较小。如果介质是由极性分子（如水）构成的，电场会使这些分子转向；这很容易做到，这种介质的 ε 就很大。在两种情形（移动质子或使极性分子转向）下，介质的极化会部分地"减弱"其内部引起的电荷（电荷⊕和⊖，图 6-12），正是如此使介质中的电场减小（与在真空中相比）。

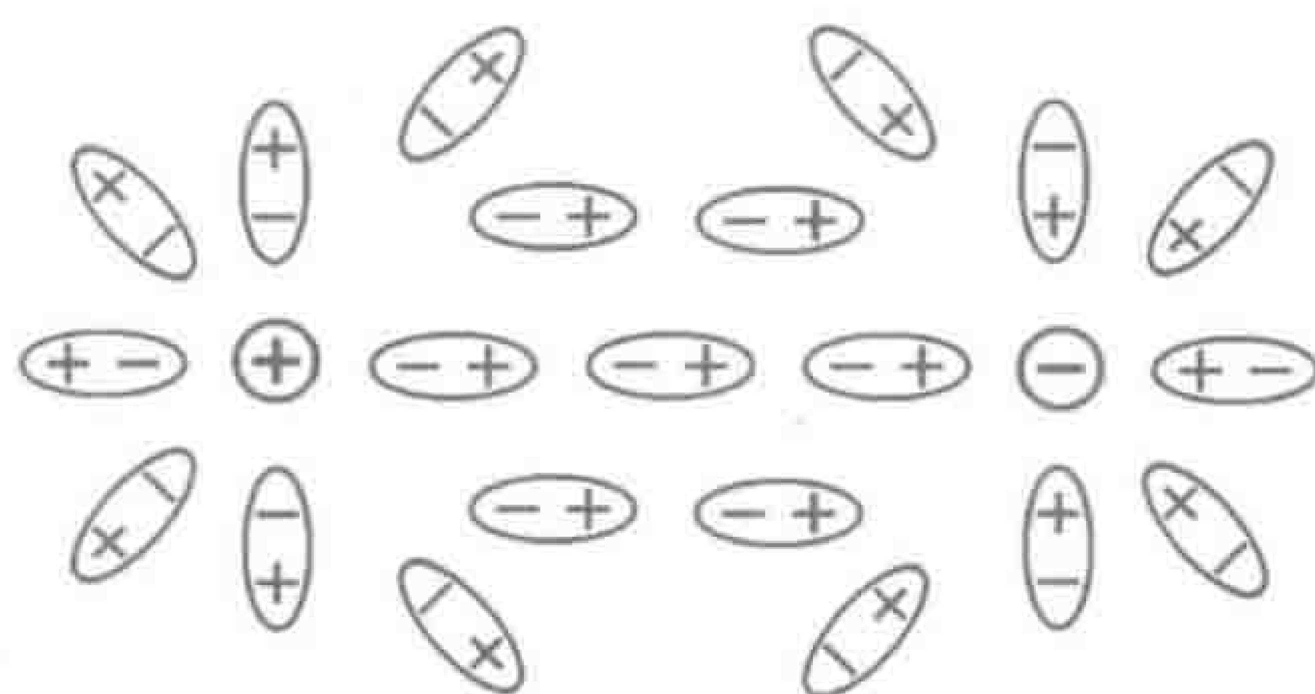


图 6-12　围绕⊕、⊖电荷和处于这些电荷之间的水分子的指向

■ 很自然地可以推测，与粒子性相关的效应应该会强烈地影响相距很近的电荷之间的相互作用：因为严格地讲，经典公式(6.1)~公式(6.3)只有在相互作用的电荷间存在大量介质分子时才能成立。而如果电荷间相距 3~4Å（在蛋白质中经常如此），任何其他原子都无法插入到两者之间并改变它们的相互作用。

看上去，在电荷这么紧密接触的情形下，它们相互作用的介电系数甚至在水环境中也应该大约为 1。这种观点（更准确些，这种误解）直到现在还能在文献中碰到。

然而，有点奇怪的是，介质的粒子性并不明显地改变介质的"宏观"（即由电荷间相距很远而导出的）介电系数，即使相距只有 3Å 左右，即甚至在这里水的介电系数值与 1 或 3 相比，更接近于 80 或 40 得多。

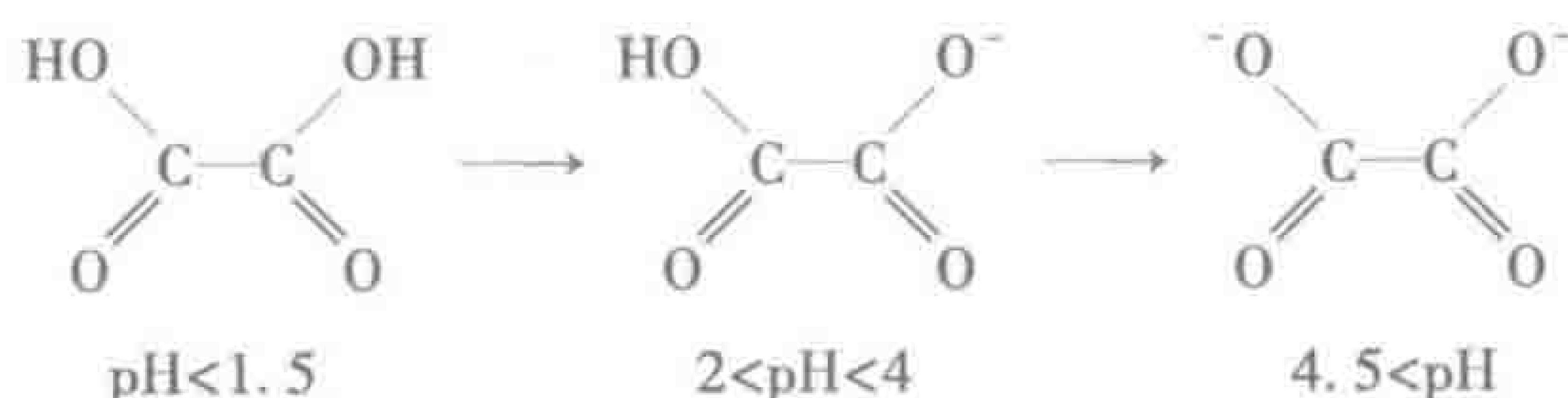
能证实这个的是，食盐能很好地溶解（离解）于水中，这只有在带电相反的离子之间在相距很近（≈3Å 时）的吸引力仍然很弱才有可能。

实际上，Na^+ 和 Cl^- 可以靠近到相距 3Å。这时它们的吸引力的自由能在 $\varepsilon=80$ 时为 −1.5kcal/mol，在 $\varepsilon=40$ 时为 −3kcal/mol，在 $\varepsilon=20$ 时为 −6kcal/mol。最后这个值（6kcal/mol）超过了氢键能量。这么大的能量会使带电相反的离子比水分子更紧地粘在一起，而饱和盐溶液浓度会是每升 10^{-4} 数量级，这与饱和水蒸气一样。但这显然不正确：把 1mol 的 NaCl（58g）溶于 1L 水中不成问题（这会是普通的、可能

①参见普通物理学中的的电磁学知识：介质在外电场中会发生极化，非极性分子主要是发生位移极化：分子的正负电荷中心出现相对位移；极性分子（如水分子）中主要是发生转向极化：等效电偶极子定向转动。

有点咸的食盐水)。意味着,甚至在3Å距离时水的介电系数也确实要大于20。

■　如果从二羟基的酸和碱在水中的一级和二级解离常数出发,可以更准确地估算在分子内部距离最短时的ε值。例如,草酸的解离过程如下:



第二级解离的pH相对于第一级解离移动了约2.5个单位,即发生于H^+浓度稀到$10^{2.5} = e^{2.3\times2.5}$倍时。这表明,第一个电荷与第二个电荷的相互作用自由能约为$2.5\times2.3RT \approx 3.5$kcal/mol。而且这是电荷间相距≈3Å时。这样的相互作用能数值对应于在相距3Å时ε≈40。在观察碳酸的解离($H_2CO_3 \longrightarrow HCO_3^- \longrightarrow CO_3^{2-}$)时得到了相似的结果:在相距≈2.5Å时有ε≈30。而"正确的"水中的氢键自由能只有在距离≈2.0Å时ε≈15~20才能得到。

当然,在距离≈1Å时ε应该接近于1,否则所有化合键相互作用的能量很难交换(这并没有被观察到!),但这种"原子内部的"距离范围更属于蛋白质化学,而不是蛋白质物理。

因此,甚至在蛋白质表面带电相反的侧链基团的**盐桥**应该只"价值"2~3kcal/mol。在蛋白质内部它应该"价值"更高,但把带电基团浸于蛋白质球体中会更贵:因此不必惊奇的是,在天然蛋白质中这种价键并不常见。

于是,我们得出结论,粒子性并不明显地改变水的"宏观"(在电荷间距离较大时导出的)介电系数,甚至在相互作用的电荷相距2~3Å时、在相互作用的电荷之间不能存在其他任何分子时也是如此。其原因是电荷被"从其他方向和侧面"靠近的介质分子(图6-13)相当强烈地屏蔽了。这些介质分子被极化了(对于水只是转向),以致这些分子的"+"朝向⊖电荷,而"−"朝向于⊕电荷。

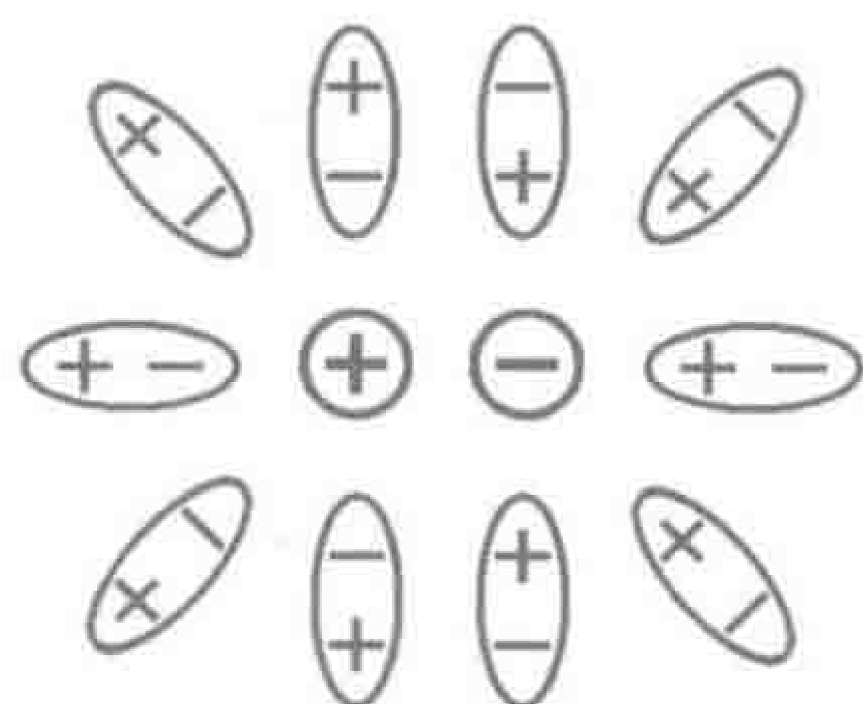


图6-13　围绕两个最大限度接近的⊕、⊖电荷的水分子的指向

在这里我们再一次(图6-10,图6-11)看到,电荷(⊕和⊖)之间的静电相互作用看上去主要是通过高介电系数的介质而产生的,弱极化的介质几乎可以忽略。

■　到此为止我们所讲到的都是"小分子"体系。这些结论对于蛋白质也成立吗?因为在那里介质(首先是水)的粒子性效应与水和蛋白质两者的介电系数巨大差异叠加在一起……

■　在Alan Fersht(蛋白质工程的首创者)的实验室里进行的实验表明,上述结论对蛋白质也完全成立。

实验原理如下。存在有某些蛋白质(酶),其活性在特定的pH下非常高(即它

们具有最适 pH)。通过基因突变向蛋白质中导入带电残基,可以使这种最适 pH(图 6-14)发生偏移,根据 pH 偏移,可以估算由突变残基在活性中心所产生的静电场。

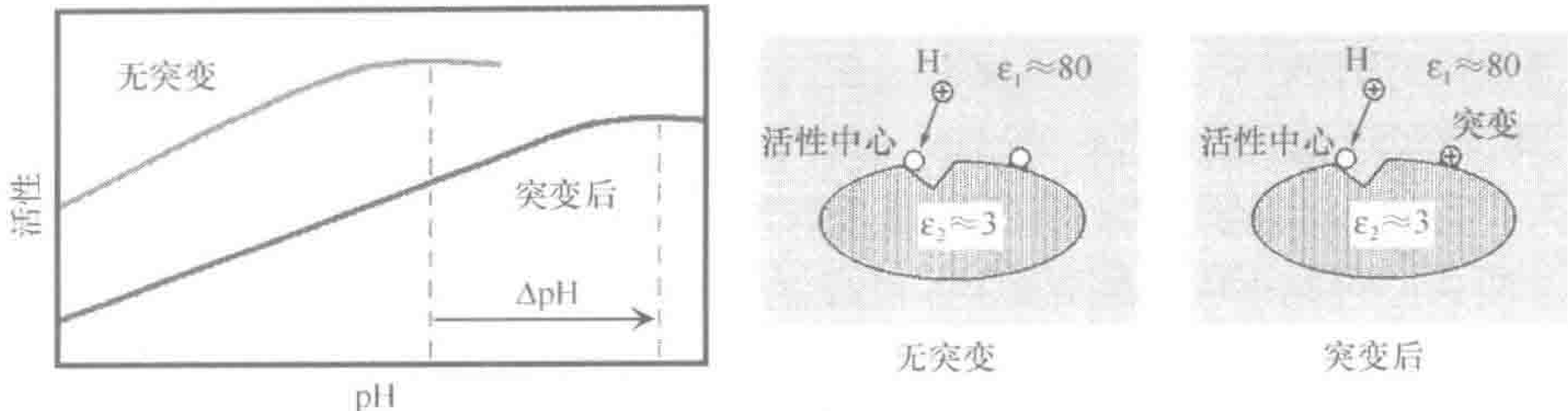


图 6-14　突变诱发的蛋白质活性最适 pH 的偏移 ΔpH。取自 Russel A. J. , Fersht A. R. 1987. Nature 328:496-500,有简化

存在最适 pH 的原因在于,为了保持酶正常工作,活性中心中的某个基团必须处于特定的带电状态,而该基团的带电性取决于介质中的氢离子浓度。H^+浓度,即 $[H^+]$(mol/L)根据定义等于 10^{-pH},而水中的 OH^-浓度约为 10^{-14+pH}。

假设活性中心(active site, AS)接受 H^+:$AS + H^+ = ASH^+$。则根据质量作用定律,相应的活性中心两种形式(带 H^+和不带 H^+)的浓度之比为

$$\frac{[ASH^+]}{[AS]} = \exp\left(\frac{-\Delta F_{ASH^+}}{RT} \cdot [H^+]\right) = \exp\left(\frac{-\Delta F_{ASH^+}}{RT} \cdot 10^{-pH}\right)$$
$$= \exp\left[-\left(\frac{\Delta F_{ASH^+}}{RT} + 2.3 \times pH\right)\right]$$

式中,ΔF_{ASH^+}为 H^+与活性中心的结合自由能(在 $[H^+] = 1mol/L$ 时),符号[ ]表示浓度。

如果突变向蛋白质中引入的电荷在其活性中心产生电势 φ,则 ΔF_{ASH^+}改变为 $\Delta F_{ASH^+}|_{有突变} = \Delta F_{ASH^+}|_{无突变} + \varphi e$,其中 e 为 H^+的电荷。因为在最适 pH 时 $[ASH^+]/[AS]$ 值(而且,相应地,$\frac{\Delta F_{ASH^+}}{RT} + 2.3 \times pH$ 的值)在无突变和突变后应该相同,则

$$\frac{\Delta F_{ASH^+}|_{无突变}}{RT} + 2.3 \times pH|_{无突变时最适值} = \frac{\Delta F_{ASH^+}|_{有突变}}{RT} + 2.3 \times pH|_{有突变时最适值}$$

即

$$\begin{aligned} \varphi e &= \Delta F_{ASH^+}|_{有突变} - \Delta F_{ASH^+}|_{无突变} \\ &= 2.3RT(pH|_{无突变时最适值} - pH|_{有突变时最适值}) \\ &= 2.3RT\ (-\Delta pH) \end{aligned}$$

因此,如果知道最适 pH 的偏移,就可以估算蛋白质突变残基在活性中心产生的电势。现在,如果知道这种蛋白质的三维结构,且相应地知道突变残基到活性中心的距离 r,则可以估算突变引入的电荷 q 与活性中心区域相互作用时的等效介电

系数 ε_{eff}[用于公式 $\varphi=q/(\varepsilon_{eff}r)$]。

在 Fersht 的实验中的突变是在蛋白质表面引入的,以便不破坏蛋白质的结构(我们已经知道,深深插入到蛋白质中的电荷的能量非常高,以至于会使蛋白质完全爆裂)。

结果:这些实验中的等效介电系数 ε_{eff}在 40~100 变动,而且前一个值是突变靠近活性中心时的,后一个值是突变远离活性中心且被蛋白质体"遮挡"时的。总体上,很大的 ε_{eff}值(尤其是它能达到 100)让人相当惊讶,因为许多对静电学不熟悉的人以为,ε_{eff}应该处于 3(如在蛋白质内部)和(最极端的情形)80(如在水中)之间。但我们对此不应该感到惊讶:它们与从图 6-10 中推导出的相当吻合。

■ 几句关于蛋白质工程的离题话。它的主要好处在于:如果所处理的蛋白质的基因、氨基酸序列及其三维结构都已知,则通过改变蛋白质基因中的某个编码,我们能够在蛋白质球体上的特定位置引入突变。除此之外,突变对结构的影响也可以监测:使用 X 射线分析或核磁共振(NMR)。因此,所有工作进行时都能看得清。

在我们所讨论的实验中,蛋白质被用来作为极微小的(更正确地说是纳米级的)静电计,蛋白质工程使得应用这种仪器成为可能,还能从物理理论跳跃到基因操作或反向进行,这是非常有趣的。

■ 现在我想对静电相互作用做几点补充。

第一点。至今为止我只讲到了单个电荷的相互作用。但是静电学还包括电偶极子(如在形成氢键时相互吸引的电偶极子 $H^{(+)}-O^{(-)}$ 和 $H^{(+)}-N^{(-)}$)的相互作用,以及电四极子[如它存在于芳香环中(图 6-15)]的相互作用。

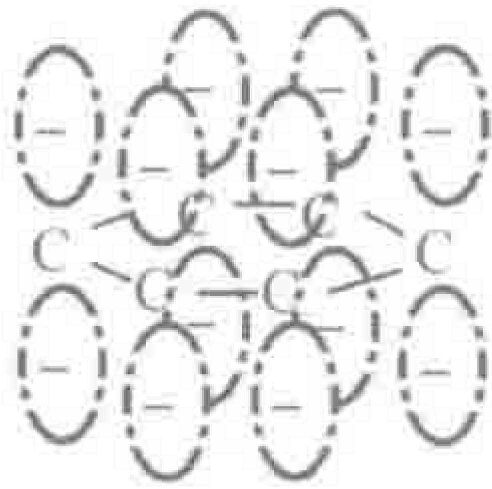


图 6-15 芳香环的电四极子:6 个 p 电子的"半"层(带电−3)(上面的云);中层为核心层(带电+6);6 个 p 电子的另外"半"层(带电−3)(下面的云)

我的讲解集中在电荷间的相互作用上,原因是它们是最强烈的:甚至在直接接触的情况下它们也比电偶极子相互作用要强几倍(而且随距离的增长而降低得更缓慢),而电偶极子的相互作用比电四极子更强。

第二点。当水中存在有自由电荷(如盐)时,静电相互作用随距离 r 的降低不是按照"能量正比于 $\frac{1}{r}$"的规律,而是快得多: $\frac{1}{r}e^{-r/D}$ 。其中 D 是德拜-休克尔(Debye-Hückel)半径:电荷周围的反离子云的特征大小。D 的值并不依赖于电荷本身,而是与介质中的盐离子的电荷和电荷浓度有关,与介质的介电系数以及温度有关。在水中,室温下,

$$D\approx 3/I^{1/2} \quad (\text{Å}) \tag{6.5}$$

其中

$$I=\frac{1}{2}\sum_i c_i z_i^2 \tag{6.6}$$

式中,I 为溶液的**离子强度**,单位为 mol/L。在式(6.6)中求和是对溶液中的所有离子种类进行的,z_i 为离子 i 的电荷(以质子电量为单位);c_i 为离子 i 的浓度。普通的生理条件下有 $I\approx 0.1\sim 0.15$mol/L;因此 $D\approx 8$Å。但是有些微生物生活在 I ~ 1mol/L 甚至更高的环境中;这时只有那些对应于"盐桥"(即电荷直接接触)的静电相互作用仍然保留,而且大大减弱了。

总的来说,在溶液中存在离子氛时,两个电荷的相互作用能量具有如下形式:

$$U=\frac{q_1 q_2}{\varepsilon_{\text{eff}} r}\times e^{-r/D}\text{。} \tag{6.7}$$

第三点。静电相互作用是粒子的**非**成对相互作用的明显例子。它不仅仅取决于电荷 q_1 和 q_2 间的距离 r,还与介质的性质(ε 和 D 都发生变化)有关,而且,尤其是与电荷到其他物体的距离及这些物体的形状有关(所有这些都影响 ε_{eff}),还与溶液中的自由离子的浓度(影响 D 值)有关。

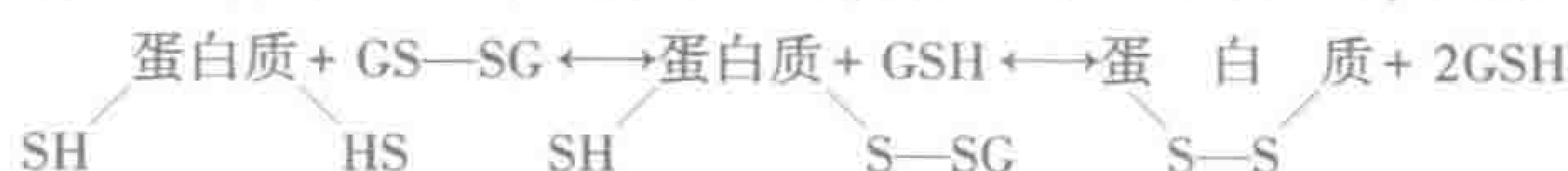
还补充一点。至今为止我通常都说"静电相互作用的能量"。这只是为讲话简便而说的:正如我在本讲开头所说的,严格地讲,应该是说"自由能"。因为我们只注意电荷的吸引和排斥,但这里并未禁止它们(以及它们的周围环境)与周围的介质**交换热量**。一旦可以交换热量,则意味着,根据定义,我们讨论的是自由能。

还有。如果研究了在水环境中静电效应对温度的依赖关系,可以证明其中的熵成分是主要的,而能量(焓)成分其实是接近于零的。这可以从水的介电系数在绝对温度 T 由 273K 变化到 373K(变化了 35%)时由 88 变到 55 (即静电相互作用增大了 40%)看得出来。而在相互作用中正比于绝对温度增长的只有熵成分,而不是焓成分。这意味着,在水中所有静电效应与能量无关,而与电荷周围的水分子有序排列有关,与电荷间相互接近或远离时这种有序排列的变化有关。

这意味着,也许看上去有点奇怪,水中的静电作用具有熵的性质,而不是能量性质,就像水环境中的疏水相互作用或氢键结合一样。

■ 在结束"蛋白质内部和周围环境的基本相互作用"这一部分时,我还想提一下二硫键和配位键。与(例如)氢键相比,这些键虽然不是很多,但是常常在蛋白质中起着重要的作用。

■ 二硫键(S—S)是由半胱氨酸(Cys)残基(侧链基团:—$C^{\beta}H_2$—SH)结合而成的(参见彩色立体图对 D-3)。伴随着水释放的半胱氨酸直接氧化作用(反应式:—CH_2—SH + HS—CH_2 ⟶ —CH_2—S—S—CH_2—+ H_2)在蛋白质中并不发生,在室温下它进行得太慢了。但是在蛋白质中,在硫醇-二硫交换反应的帮助下,二硫键的形成可以快速进行。有人认为,在细胞中这是在**谷胱甘肽**(glutathione)[它具有硫醇单体(GSH)和二硫二聚体(GSSH)两种形式]的参与下发生的,且按下式进行:



在细胞中，二硫键的断裂和形成都是由特异性的酶：二硫异构酶所催化的（即加速，但并不改变方向）。

在细胞中二硫键的形成是可逆的，因为硫醇-二硫交换反应的能量平衡接近于零（反应前有两个 S—H 共价键和一个 S—S 键，反应后也是这么多；这很像在水环境中蛋白质内部形成氢键时的能量平衡，不是吗?）。而且，细胞中 GSH 有效浓度相当高，会使平衡向使在半胱氨酸“偶然”接近时可能形成的那些键发生断裂的方向移动。因此在这里能够形成并维持的二硫键只能存在于已经靠近（在其他相互作用下）的半胱氨酸之间。

二硫键对于驻留在细胞外工作的蛋白质非常重要。一方面，细胞外没有二硫异构酶和谷胱甘肽，因此已经结合的（在细胞内或在离开细胞时）键被“冻结”，它们不会断裂或改建。另一方面，胞外的蛋白质会进入到不同的环境中，稳定的冻结二硫键会给蛋白质额外的强度储备，使它不会被干扰。因此，与胞内蛋白质相比，二硫键对分泌性蛋白质尤其典型。通常，在分泌蛋白质中所有可用的半胱氨酸（如果其数目为奇数，则剩余一个）都参与形成二硫键。

■ **配位键**是由蛋白质的 N、O 和 S 原子（以及水的 O 原子）与二价或三价金属离子（Fe、Zn、Co、Ca、Mg 等）形成的。

这些金属离子在仅比它们的满电子轨道略高一点的低能级处有未填满的电子轨道。每条空轨道都能结合电子对。而 O、N、S 原子（“电子供体”）带有电子对，能够占据离子的空轨道。在这种情形下形成的键与普通共价键没有区别，只是在普通键中电子是由成键的两个原子提供的，而配位键中只由其中的一个原子提供电子。

在形成配位键时，金属离子与几个电子供体相结合。于是几个大的（半径 ~1.5Å）供体从各个方向包围小的（半径 ~0.7Å）二价或三价离子。最常见的是 6 个供体原子包围（配位）金属离子，各自占据正八面体的顶点（图 6-16）。

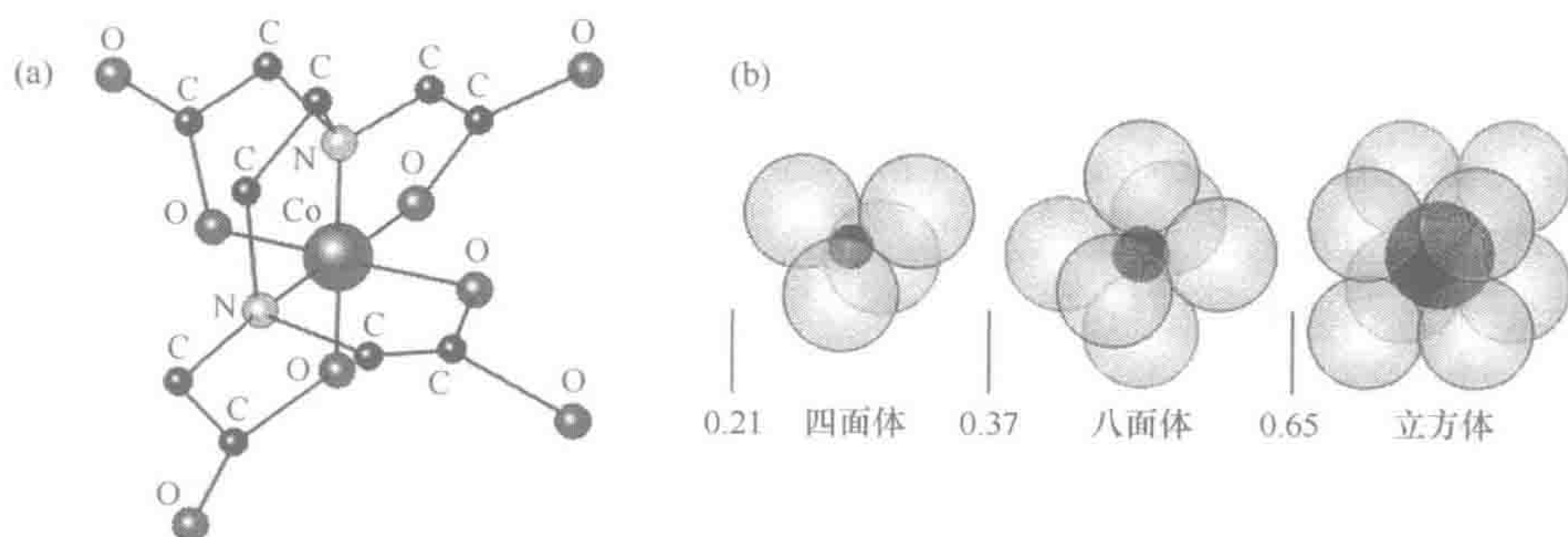


图 6-16 (a) Co^{3+} 与 EDTA（乙二胺四乙酸）形成的八面体复合物的结构；(b) 中心离子的半径与围绕它的电子供体半径不同比率下，中心离子的典型配体情况。图片取自文献[10]，有改编

因为离子可以与蛋白质电子供体相结合，也可以与水的氧原子结合，它（尽管每个键的能量很高）可以在水和蛋白质之间互相转移，而没有巨大的能量增加或损失。看来，更重要的是另一方面：如果在蛋白质中供体原子已经处于“正确的”（就形成配位键而言）位置，则离子可以放开原先环绕在它周围的水分子，而与蛋白质中的原子结合。由于释放的水分子获得了运动熵，所形成的键更加牢固。平均而言，每个配位键价值几个 kcal/mol 的能量，比水中的氢键要高一些。

由一个分子中的几个原子与单个离子结合成的这些键被称为**螯合的**（“爪形的”）。螯合键在蛋白质（尤其是在它们的活性中心）的作用，我们将在后面讨论。我们还将看到，完全包被离子的螯合复合体可以进入到蛋白质的疏水核。而现在我想再一次让大家的注意力转到图 6-16 上，该图中展示了在实验室中经常使用的试剂 EDTA（乙二胺四乙酸）①与金属的螯合。由于 EDTA 的 COO^-基团的负电荷与带正电的金属离子相结合，螯合键非常牢固。

①EDTA 是一个六齿配体，可以螯着多种金属离子。它的 4 个酸和 2 个胺的部分都可作为配体的齿，与锰、铜、铁及钴等金属离子组成螯合物。它是滴定中常用的螯合剂，往往用于测定金属离子的浓度，精确而敏感。

多肽链的二级结构

第 7 讲

多肽的二级结构。螺旋：2_7，3_{10}，α，π，poly(Pro) II。反平行和平行 β 结构，β 转折。测定二级结构的实验方法。

在学习了基本相互作用之后，今天我们来讨论蛋白质的二级结构。首先我们来看规则的二级结构：α 螺旋和 β 结构，但不仅限于它们。

α 螺旋和 β 结构折叠成球，决定了蛋白质的三级结构(图 7-1)。这些二级结构的区别在于主链的特定周期性构象，而侧链构象各不相同。

首先来看螺旋。它们可能是左旋的或右旋的(图 7-2)，它们可以有不同的周期和螺距。右旋的(R)螺旋逆时针方向(相应于几何学中的正角)转动时趋向我们；而左旋的(L)则以顺时针方向转动时趋向我们。

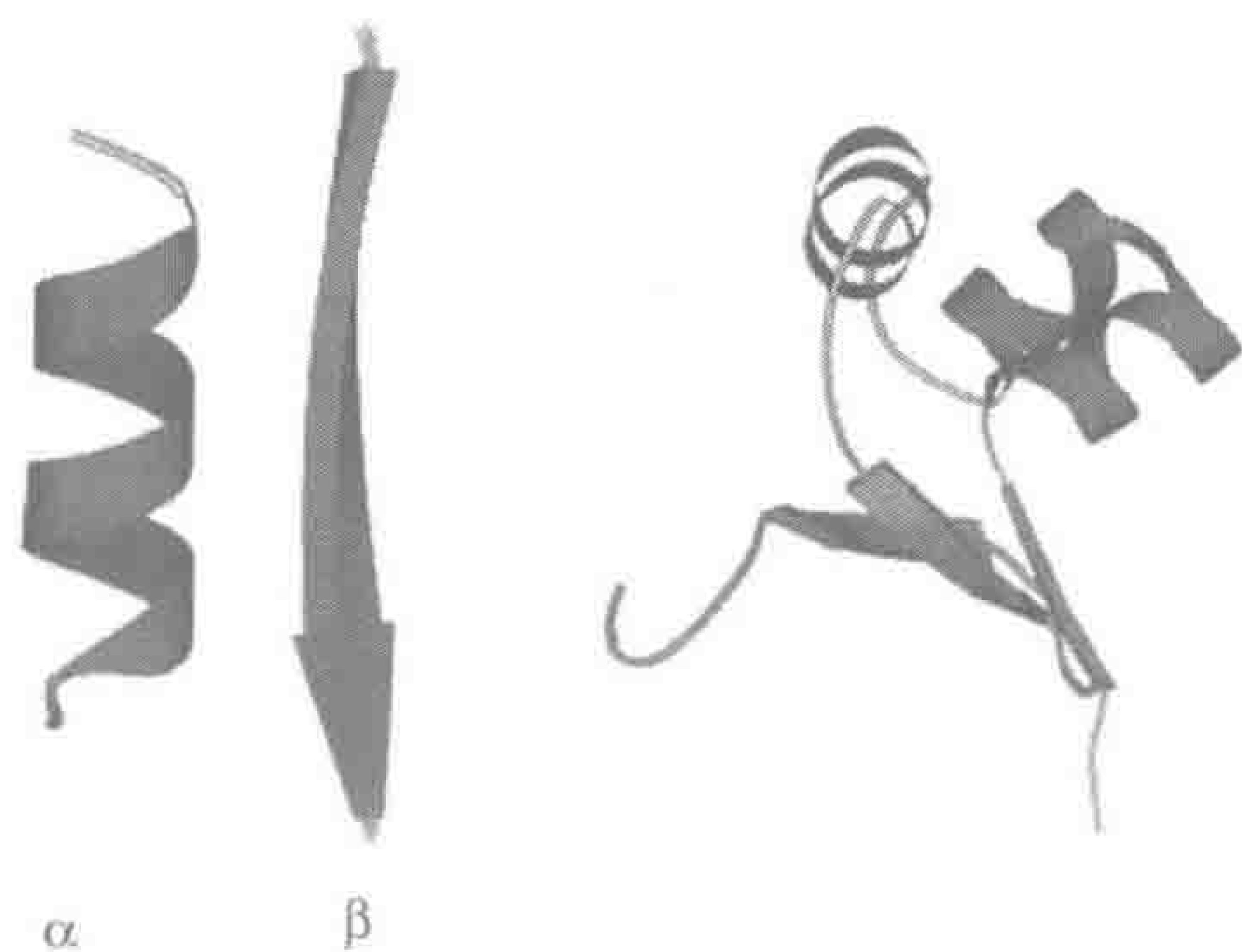


图 7-1　多肽链的二级结构(α 螺旋和一条 β 片)和蛋白质球的三级结构。通常，α 和 β 结构加起来约占球蛋白链的一半

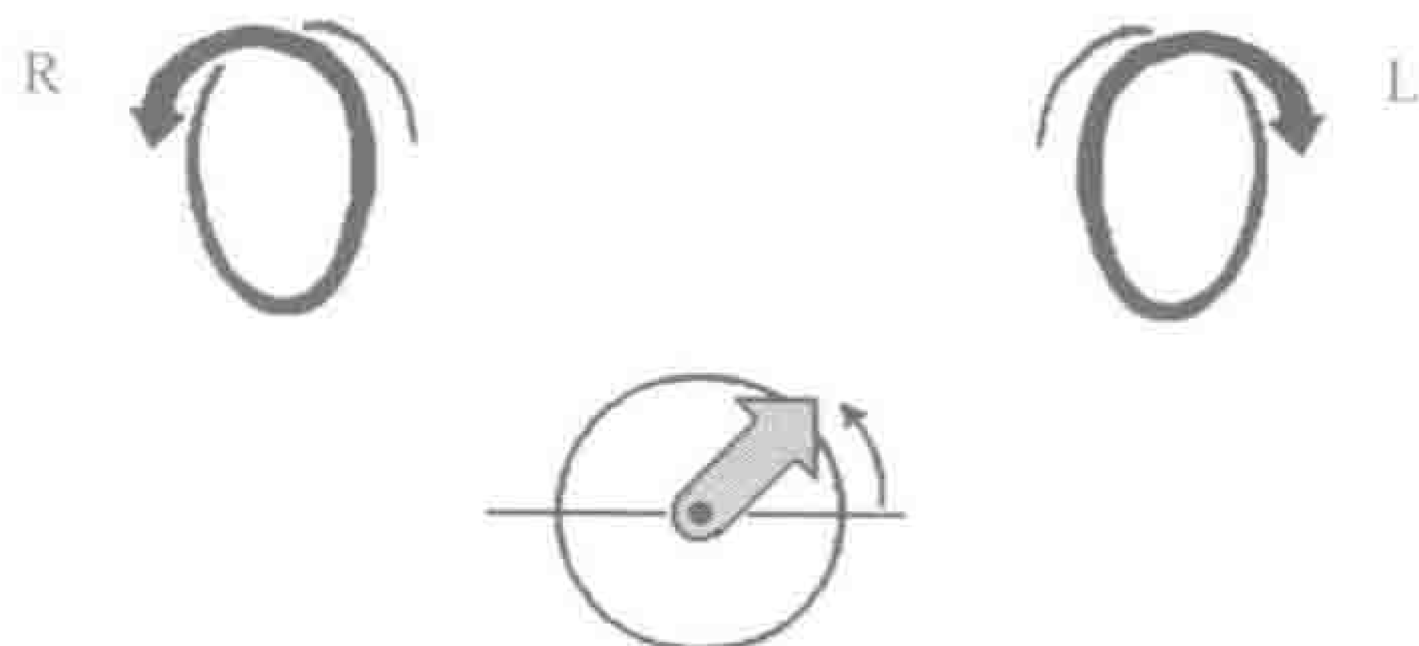


图 7-2　左旋螺旋(R)和右旋螺旋(L)。下方显示了三角函数的正角读数：其中逆时针旋转的箭头"趋近我们"。当然，在螺旋(图 7-5，参见彩图立体图对 D-4)中是可从任意一端沿着它的轴看过去

在肽链中最重要的螺旋是由氢键支撑的，肽骨架上的 C ═O 基团与位于它们

C′ 端链方向的 N ═H 基团结合。原则上,由氢链拉紧的螺旋可以有以下几种(图 7-3):2_7,3_{10},4_{13}(通常称为 α)和 5_{16}(通常称为 π)。在名称“2_7”中,“2”表示链上的第二个残基(图 7-3),“7”表示这个氢键连接而成的环(O···H—N—C′—C^α—N—C′)上的原子编号。其他螺旋的名称中的数字意义与此相同。

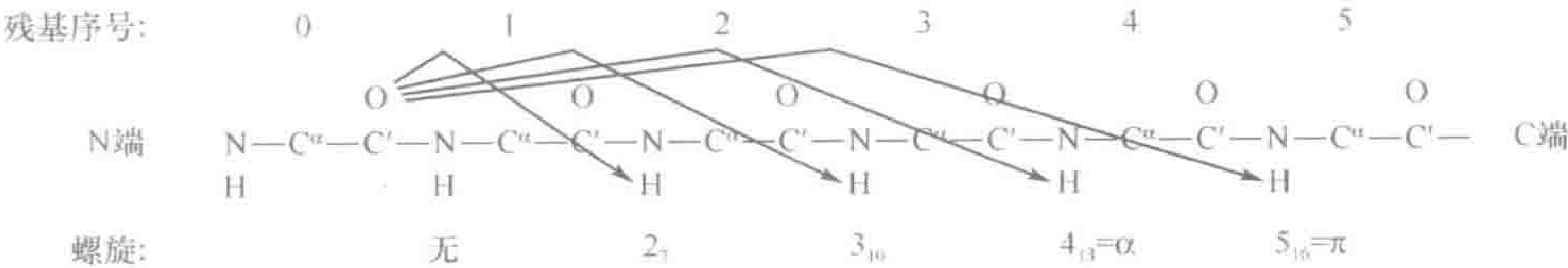


图 7-3 不同螺旋的典型氢键(以箭头显示)

■ 这些螺旋结构中,哪一种在蛋白质中占优势呢? α 螺旋。为什么? 典型氨基酸残基,丙氨酸(Ala)的拉氏图(图 7-4)给出了这个问题的答案。在图中标出了一些构象,其周期性重复导致了图 7-3 中所示的氢键连接。

可以看出,只有 α_R(α 螺旋)可以深入到丙氨酸(以及所有其他残基)允许区的内部。其他螺旋(如左旋的 α_L 或右旋的 3_{10})要么位于这个区域的边缘(这些位置的构象张力已经增大了);要么处于仅甘氨酸允许的区域。

因此可以预料的是,仅仅只有右旋的 α 螺旋应该是最稳定的,事实上正是如此,因此在蛋白质中占据优势,正如所观察到的那样。在右旋 α 螺旋(图 7-5,参见彩色立体图对 D-4)中所有原子的排列是最优的:紧密而无张力;因此不难理解,在蛋白质中这种螺旋很多,而在纤维蛋白中它们能延伸到极大的长度,并包含数百个氨基酸残基。

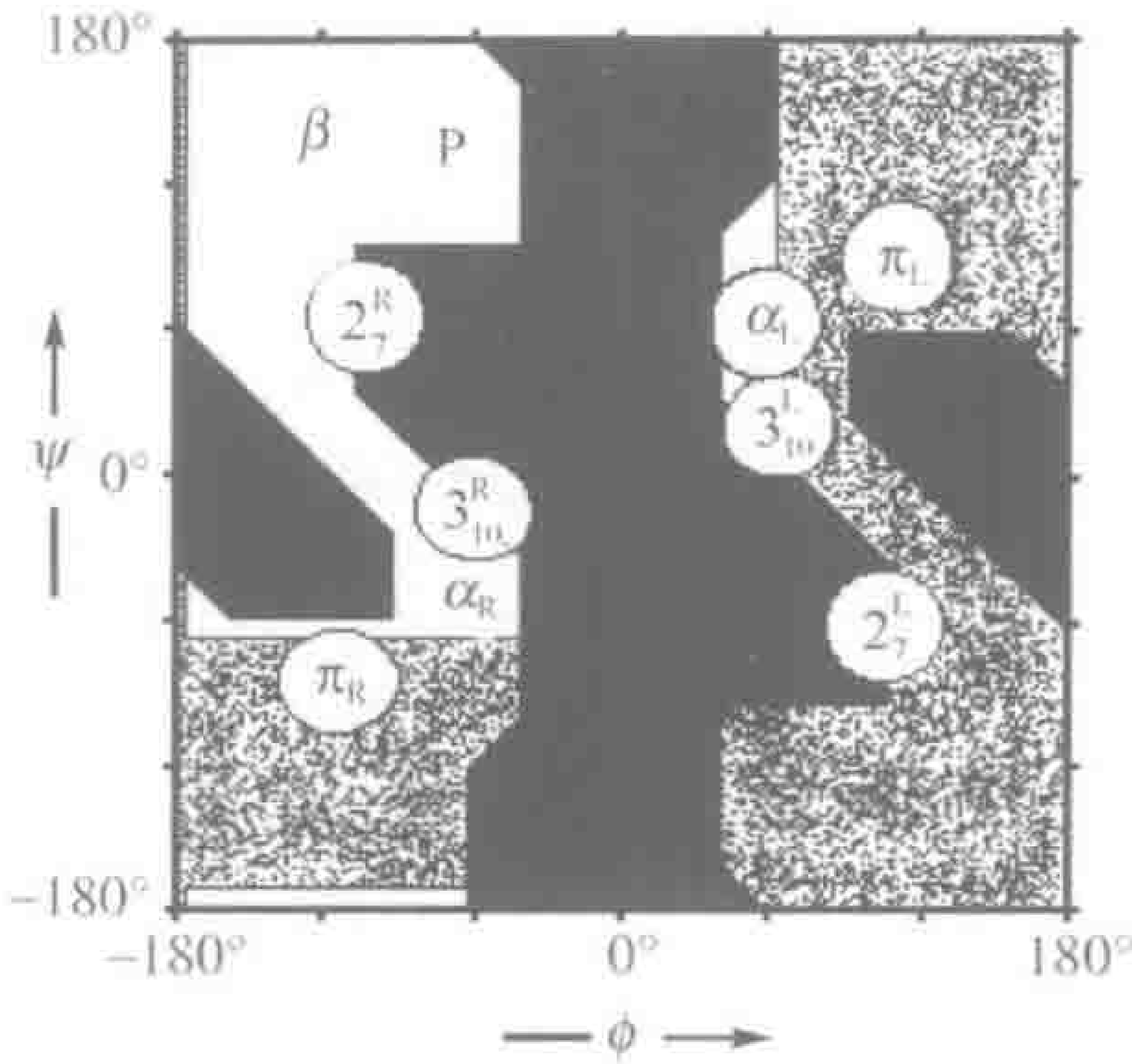


图 7-4 在氨基酸残基的允许和禁止构象图背景上的不同二级结构的构象。2_7^R、2_7^L:右旋和左旋的 2_7 螺旋;3_{10}^R、3_{10}^L:右旋和右旋的 3_{10}螺旋;α_R、α_L:右旋和左旋的 α 螺旋;π_R、π_L:右旋和左旋的 π 螺旋;β:β 结构[详见图 7-8(c)];P:多聚脯氨酸[poly(Pro)]II 型螺旋。□:丙氨酸(Ala)允许的构象区域;▒:仅甘氨酸(Gly)允许而丙氨酸和其他残基禁止的构象区域;■:所有残基都禁止的区域。φ 和 ψ:主链中的内部转动角

左旋 α 螺旋在蛋白质中实际上不存在。2_7 螺旋也不存在,它不仅处于允许区的最边缘,而且 N—H 基团与 O ═C 基团之间的夹角几乎是直角,对氢键是能量不利的。

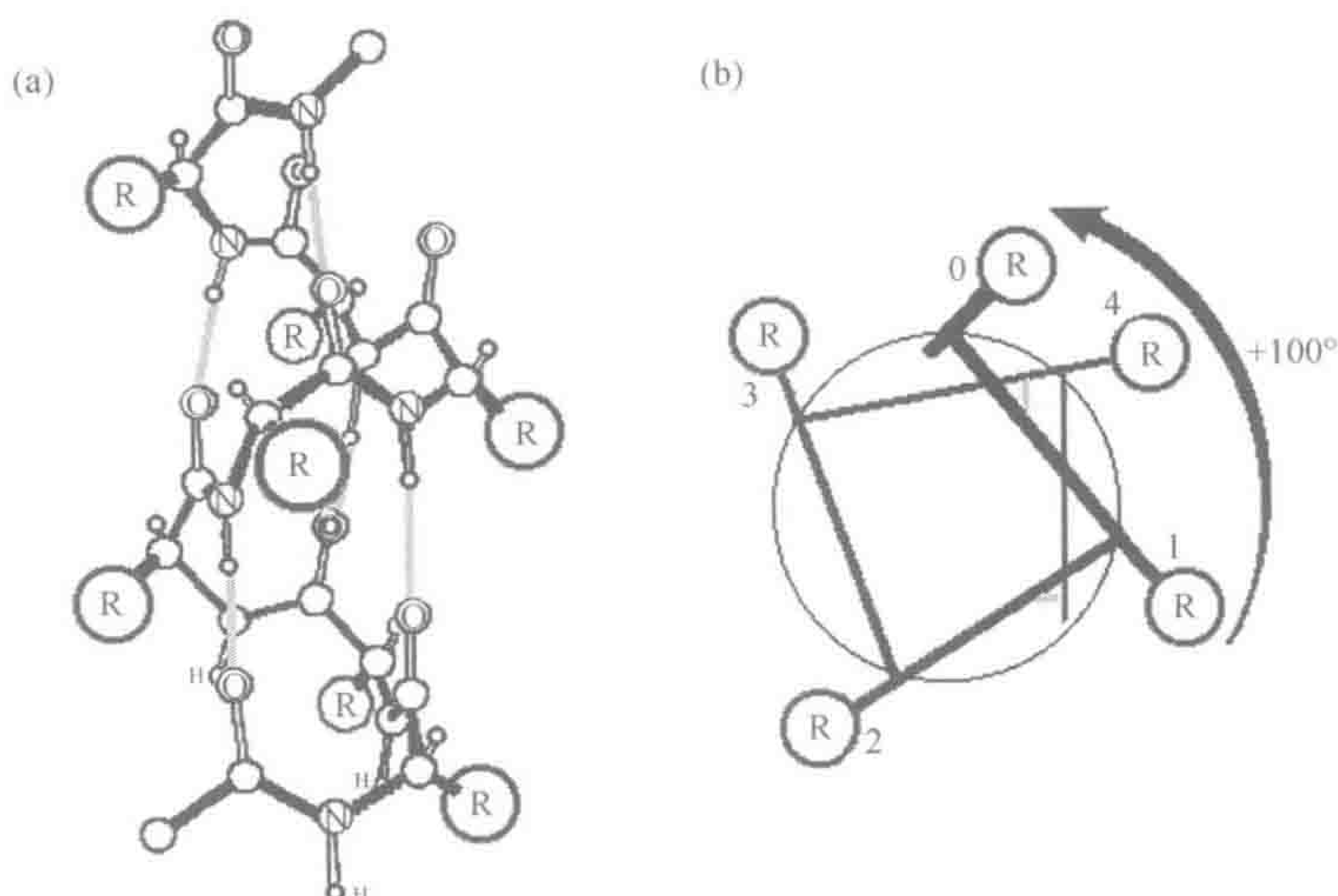


图 7-5　右手 α 螺旋。(a) 原子结构。R：侧链基团。灰线：氢键。(b) 此 α 螺旋转动一圈的示意图，从顶部往下看。箭头表示螺旋趋近我们(残基编号按接近读者而减小)的转动方向(以单个残基计)。图片取自文献[3]，有改编

在蛋白质中实际上也不存在 π 螺旋。它们也处于允许区的最边缘，而且它们的螺圈太大，导致它们在轴上有一个能量不利的"空洞"。而 3_{10} 螺旋[主要是右旋的，左旋的实际上只有甘氨酸(Gly)存在]在蛋白质中存在，不过只是以短的(由 3~4 个残基组成)扭曲变形片段的形式(因为有空间张力：螺旋拉得太紧！3_{10} 螺旋相应的构象处于允许区的最边缘)。

注意在图 7-5(a)中可以清楚地看出螺旋的一个性质：在它们的 N 端有未参与螺旋内部氢键的 N—H 基团的 H 原子，而在 C 端，有未参与氢键的 C ═O 基团的 O 原子。由于 H 原子的电子云部分地被带负电的 N 原子所拉开，而电负性的 O 原子自己吸引 C′ 原子的电子，使 N 端带有部分正电荷，而 C 端带有部分负电荷。因此，螺旋就成为一个长电偶极子：在 α 螺旋 N 端的总的"+"电荷(由三个 NH 基团所提供)大约为半个质子电荷，而 C 端的"-"电荷大约为半个电子电荷。

■　现在来看在内部没有氢键、但彼此之间有氢键结合的规则结构。

伸展的[主链中的所有角几乎都是反式(*trans*)]、略微扭转的肽链形成 β 结构。它(图 7-6)有平行型(β↑↑)、反平行型(β↓↑)和混合型(由 β↑↑和β↓↑组成)。β 结构是由氢键拉紧的(图 7-6 中的/ \| 符号)。它以或多或少的大块薄片形式存在。由于 β 结构的表面是凸凹不平的，也被称为"β-褶皱结构"(图 7-7)。

β 折叠片在总体上总是有点扭转[图 7-8(a)，参见彩色立体图对 D-5~D-7]，原因在于单个 β 串是稍微扭曲的，从而沿着 β 串走向上氢键的方向稍微有所改变。而串的扭转，正如 C. Chothia 所解释的，是带有侧链基团的残基的能量最有利构象

向立体空间允许区域中心偏移所致[图 7-8(c)]。单个 β 串的扭曲是左旋的(在 L 型氨基酸中！在 D 型中会是相反的)：图 7-8(b)中可见，当 β 串接近我们时，侧链基团是顺时针方向转动的(每个残基转动约-165°)。

β 串扭转的结果是氢键也发生转动(每个残基转过-165°，即每对残基 β 结构的重复单位转动-330°= +30°)，因此 β 结构相邻串之间的角度[如果从折叠片边缘看过去，图 7-8(a)]通常约为-25° (“负”总是意味着 β 串相对于较远的 β 串按顺时针方向转动接近我们)。所以，如果沿该片的边缘看过去，β 折叠片是**左**旋的(如果沿 β 串按氢键线的转折看过去，是**右**旋的)。

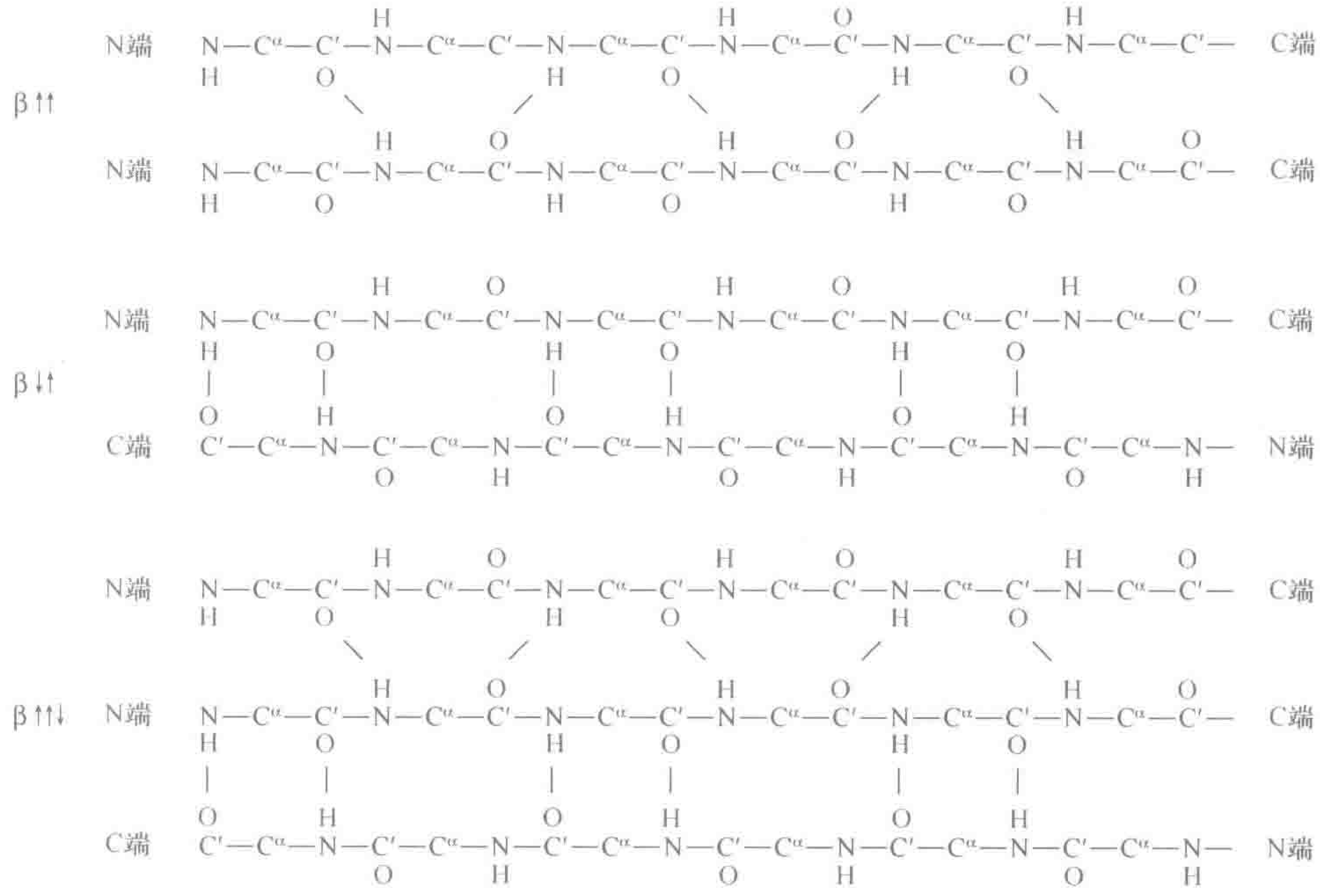


图 7-6 在平行型 β↑↑、反平行型 β↓↑ 和混合型 β↑↑↓ 的 β 结构中的肽链走向和氢键分布示意图。可见，β 串中一个残基的氢键朝向一个方向，而下一个朝向另一个方向，如此交替

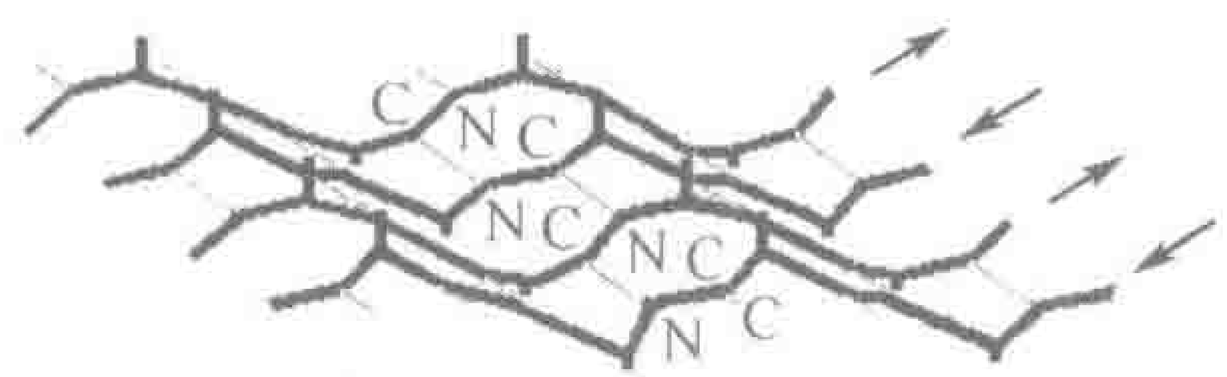


图 7-7 β 片结构具有褶皱表面。细灰线表示 β 串中的氢键。侧链基团(见突出短线)分布在褶皱上；每一个都朝向与褶皱相同的方向，即沿 β 串的侧链基团轮流朝下和朝上。图片取自文献[3]，有改编

■ 存在有不含氢键的螺旋。它们的紧密(意味着：能量有利的)排列仅由范德华力维系。这就是多聚脯氨酸螺旋。它由三条链组成，每条链绞合成为颇为伸展的

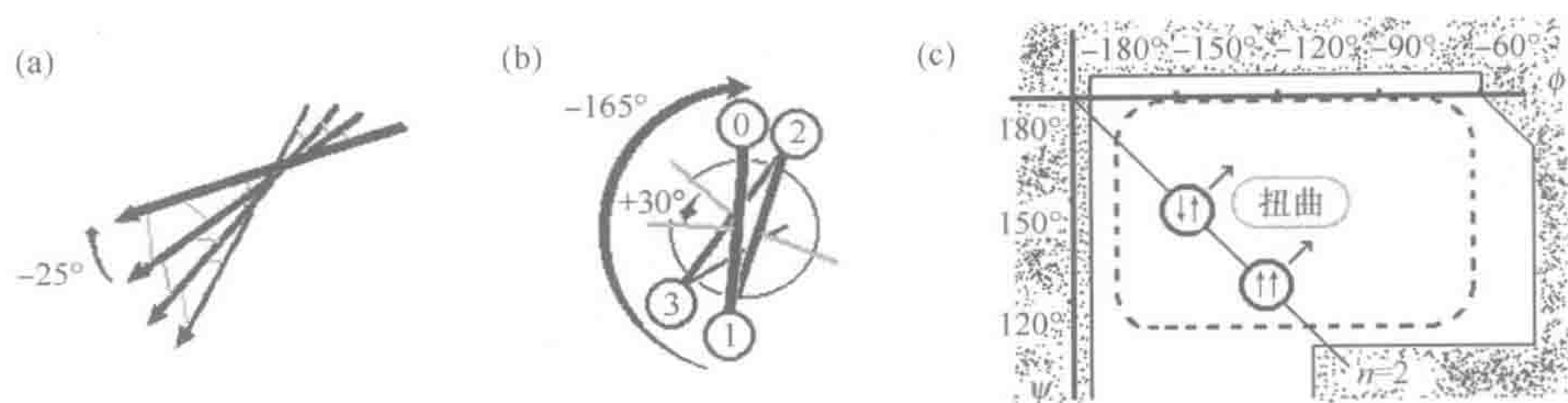


图 7-8 (a)β 片的扭转。β 串显示为箭头,其中的氢键显示为细灰线。(b)β 串转一圈的示意图,俯视图。小圆圈为侧链基团,其编号随远离读者而增加。细灰线指示在 β 片中参与氢键的 C ═O 基团的方向。大箭头:β 串趋近读者时单个残基的转动;小箭头:β 串趋近读者两个残基时同向氢键的指向转动。(c)理想的(无扭曲)聚合甘氨酸[poly(Gly)]平行(↑↑)、反平行(↑↓)β 结构构象和实际的(由 L 氨基酸组成)扭曲 β 结构的平均构象。虚线表示单个丙氨酸(Ala)残基的能量最小值区域;阴影界线显示了该残基的允许构象区的边界。$\phi\psi$ 图的对角线对应于每圈 2 个残基的扁平周期结构。对角线之上为左旋(L)螺旋,下方为右旋(R)螺旋。(a)和(c)取自文献[3],有改编

左旋螺旋。它们紧密地相互缠绕,形成右旋的超螺旋。在多聚脯氨酸螺旋的两种可能类型,对我们比较重要的是Ⅱ型[poly(Pro)Ⅱ]:这种螺旋能在胶原中形成。在这种螺旋中,脯氨酸肽基处于普通(反式)构象。在本课程相应的地方我们会更仔细地研究胶原螺旋,在此我们仅限于介绍其外观(图 7-9),请注意在图 7-4 中其肽链构象所对应的区域:可见,它相当接近 β 结构。



图 7-9 由三个左旋螺旋构成的多聚脯氨酸Ⅱ型[poly(Pro)Ⅱ]超螺旋的总体外观

最重要的几种蛋白质链规则二级结构的参数总结于表 7-1 中。

表 7-1 分布最广的几种蛋白质二级结构的基本几何参数

结构	氢键	每圈残基数	每个残基偏移量/Å	ϕ	ψ
α_R 螺旋	CO_0—HN_{+4}	+3.6	1.5	−60°	−45°
$(3_{10})_R$ 螺旋	CO_0—HN_{+3}	+3.0	2.0	−50°	−25°
β↑↓ 折叠片	肽链之间*	−2.3	3.4	−135°	+150°
β↑↑ 折叠片	肽链之间*	−2.3	3.2	−120°	+135°
poly(Pro) Ⅱ螺旋	无	−3.0	3.0	−80°	+155°

*β 折叠片中串间距离:4.8Å;

注:数据取自文献[3],文献[6]。所有数字舍去尾数。“每圈残基数”列中的“+”表示右旋螺旋,“−”表示左旋。

■ 除了规则二级结构之外，在肽链中还有非规则的二级结构，即不形成长条周期系统的标准结构。

这就是所谓的 β 转折（β-turn）（"β"是因为它们经常是连接反平行 β 发夹中的相邻 β 串尖端）。最重要的 β 转折的典型外观和其组成残基的构象显示于图 7-10 中（参见彩色立体图对 D-8～D-10）。请比较图 7-10（c）与图 7-4 和表 7-1，注意 I 型（特别是 III 型）转折的构象与 3_{10}螺旋圈的构象很相似。

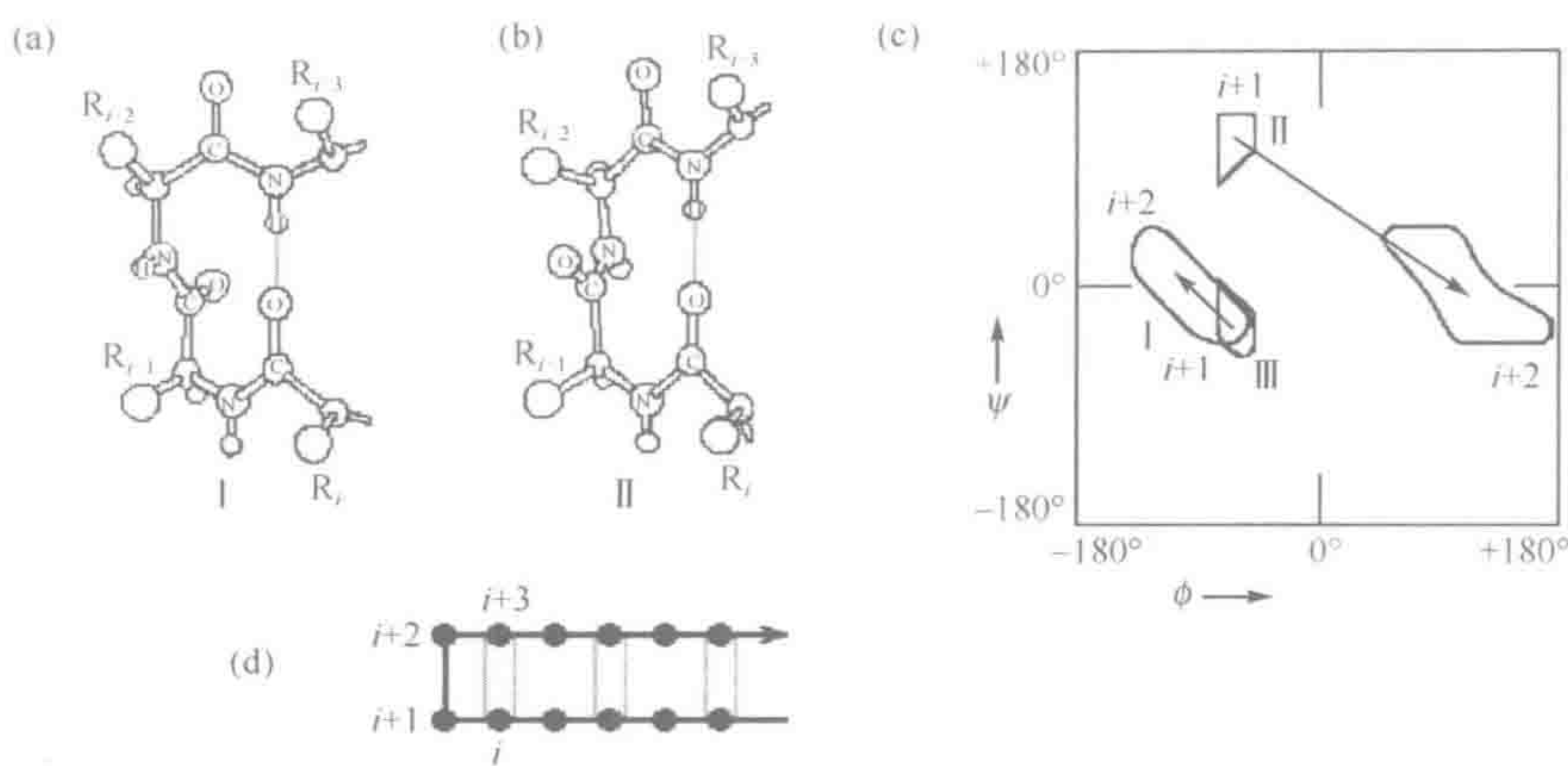


图 7-10 β 转折。（a）I 型 β 转折（III 型 β 转折与 I 型非常相似，因此未单独绘出）。（b）II 型 β 转折。它与 I 型的主要差别在于相连的残基 i+1 和 i+2 的肽基团是翻转的。（c）β 转折中由氢键所固定的 i+1 和 i+2 残基的构象。在 III 型 β 转折中残基 i+1 和 i+2 具有相同的构象（粗点表示）。β 转折中残基 i 和 i+3 的构象并不固定；如果 β 结构如图（d）（顶端带有 β 转折的 β 发夹示意图）所示，由转折向外伸长，则残基 i 和 i+3 的构象由 β 结构所固定。（a）、（b）和（c）取自文献[3]，有修改

未包含于蛋白质规则二级结构中的残基大致有一半处于转折中。

■ 在本讲结尾处，就实验中如何观测二级结构讲几句。

无疑，如果对蛋白质做了 X 射线（或者精确多维核磁共振）实验，则二级结构可以根据原子坐标推导出来。

然而，由于能够检测靠得很近（4～5Å 或更小）的原子核，甚至在尚未获得完整的蛋白质原子结构之前，核磁共振（nuclear magnetic resonance，NMR）都能够确定二级结构。

NMR 方法基于使用射频波激发置于强磁场中的原子核。这些原子核的核子（质子和中子）数必须为奇数：于是它们有自旋，从而有磁矩。在蛋白质中，这类原子核有天然的"轻"氢^{1}H 和引入的同位素^{13}C、^{15}N 等。磁共振发生于被测原子的特征射频波长，且该波长会被该原子的共价键相连和空间相近的邻近原子稍微改变（从而有助于判断被激发的是哪个残基的哪个原子）。激发的原子核可以把激发状态传递到空间上相邻的有磁矩的原子核，接受者会在自己的频率上报告获得的

激发情况(从而可以判断这两个磁性核的靠近程度)。

α 螺旋的特征在于 $C^{\alpha}H$ 基团的氢原子与第 4 个(往 C 端方向)残基 NH 基团的氢原子相靠近,而 β 结构的特征是 NH 基团的氢原子和链上直接相连或 β 折叠片中由氢链相连的残基的 $C^{\alpha}H$ 基团的氢原子相靠近(图 7-11)。

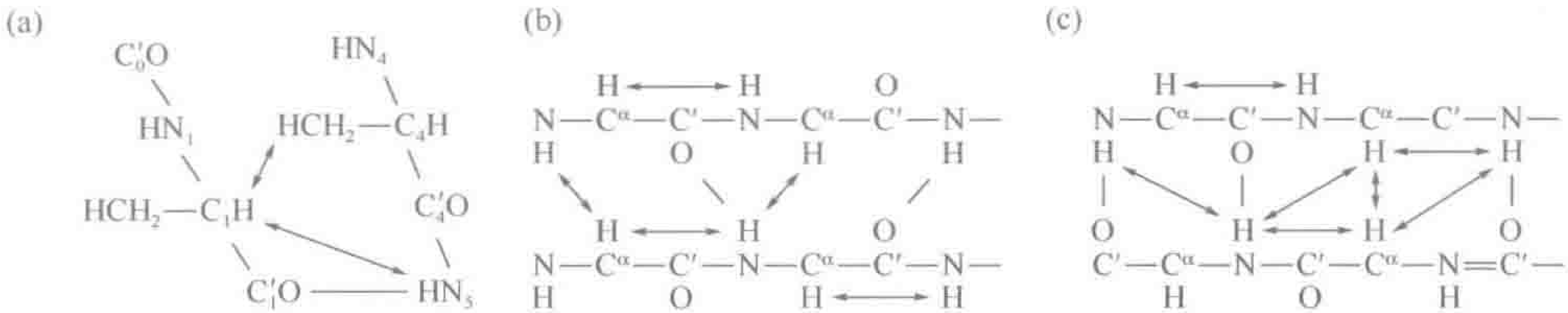


图 7-11 α 螺旋(a)、平行 β 结构(b)和反平行 β 结构(c)中的最典型的氢原子核的靠近程度(↔)。在图(a)中主链原子的标号显示了链中残基的相互位置

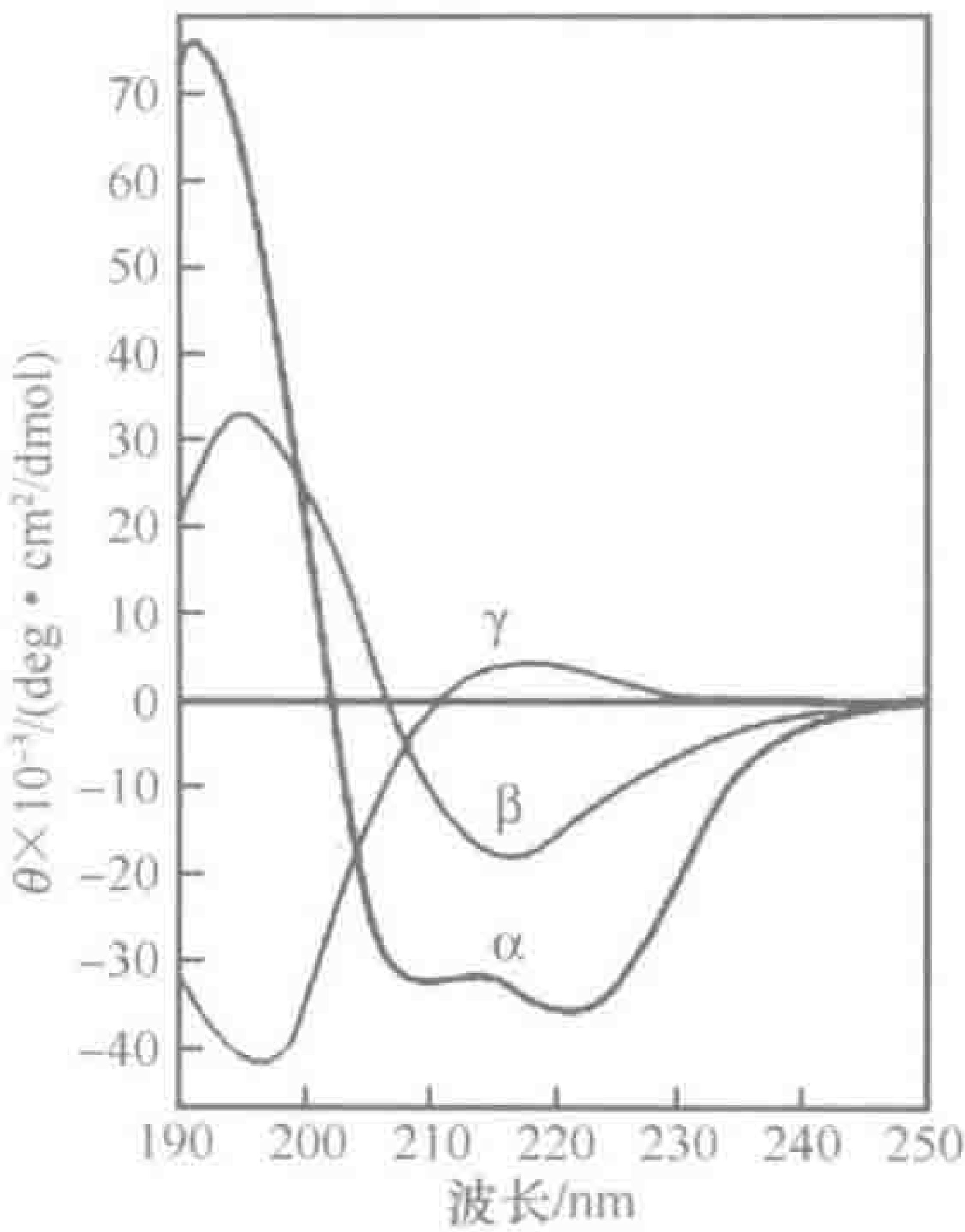


图 7-12 聚赖氨酸(polylysine)在 α 螺旋(α)、β 结构(β)和无规线团(γ)形态中的典型圆二色性谱。图片取自 Greenfeld N. J., et al. 1969. Biochemistry 8: 4104-4116,有修改

不过,在测定二级结构方面更重要的方法是圆二色性(circular dichroism, CD)。它不需要蛋白质的基本空间结构知识。相反,蛋白质的结构研究通常是从获得圆二色性谱开始的。圆二色性方法的基础是不同扭转度的螺旋中对右旋和左旋偏振光的吸收度有差别。由于这种吸收度差别,平面偏振光变成了椭圆偏振光。

在"远"紫外区(190~240nm)的典型圆二色性谱如图 7-12 所示。所显示的谱取决于肽基团周围环境的对称性,因此反映了在蛋白质中是否有二级结构,是哪种,有多少。

在远紫外区,肽基团在波长约为 200nm 处发生光激发。这大致是单个原子激发所需波长的两倍。肽基团能被波长更长(即不大"硬")的光所激发,原因在于肽基团上的电子会按几个原子发生离域化,我们前面已经讲过。

在芳香族侧链基团中,电子离域更大:它们不像在肽基团中那样只"传播"3 个原子,而是 6 个原子。芳香族侧链基团的圆二色性谱发生在波长 250~280nm 处(虽然"尾部"可达~220nm)。在这个 250~280nm 波长区域("近"紫外区)可研究芳香族侧链基团环境的不对称性,即与形成蛋白质三级(而不是二级)结构有关的效应。

顺便说一句,当电子在更大范围内(含多价键的更大分子中)离域化时,它已经不是由紫外线,而是由可见光开始激发:这些分子的发光可由眼睛观察到,即它们可成为染色剂。

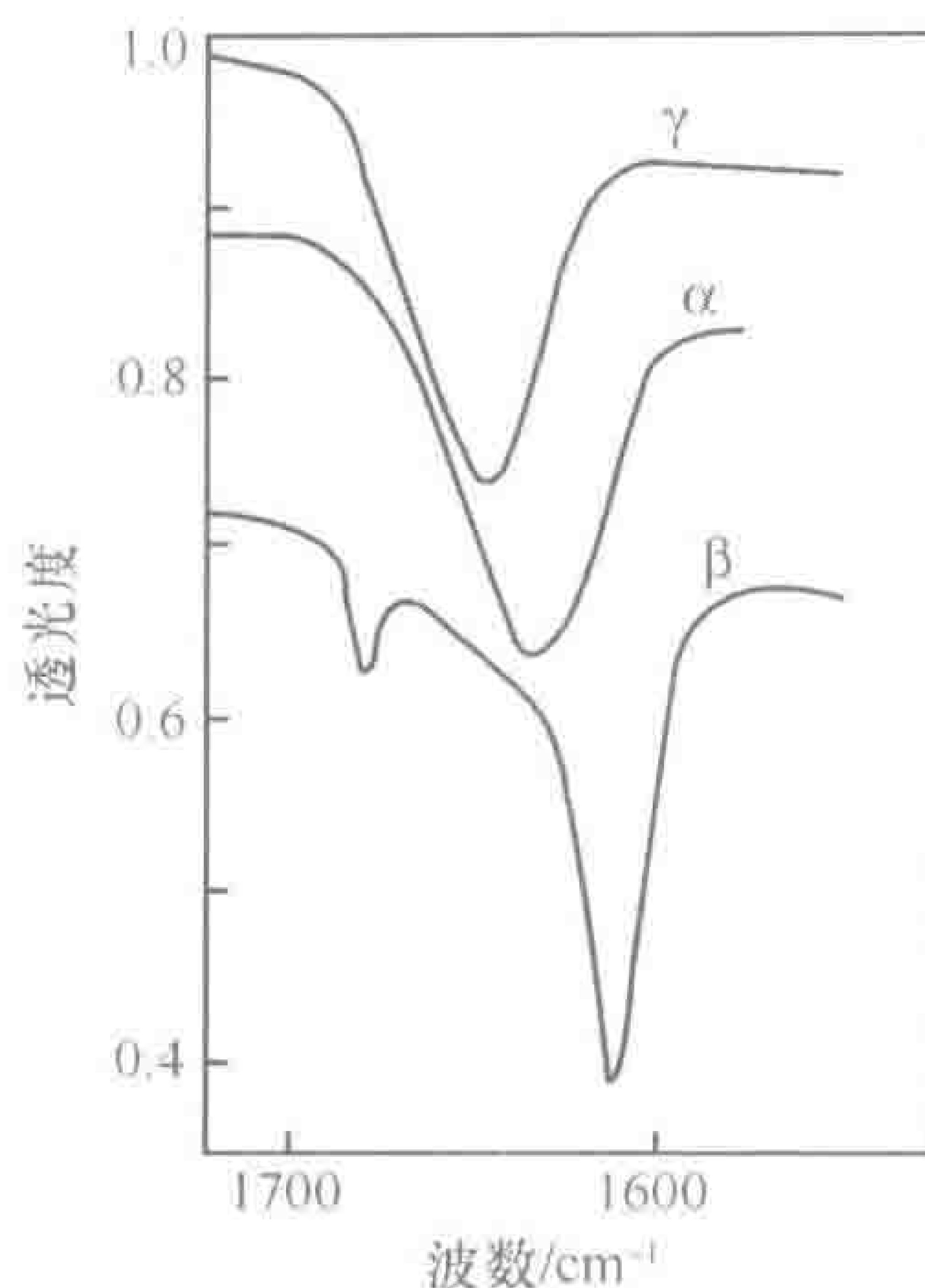


图 7-13　在重水(D_2O)中测得的聚赖氨酸(polylysine)在 α 螺旋(α)、β 结构(β)和无规线团(γ)形态下的典型红外透射光谱。测量是在反映 C ═O 连接的"酰胺 I"区域中进行的。图片取自 Susi H. 1972. Mothods Enzymol. 26:455-472,有修改

除了紫外光谱之外,测量多肽和蛋白质的二级结构还可使用红外光谱。它们反映了参与和未参与不同二级结构的肽基之间在振动上的差异(图 7-13)。红外光谱的测量比紫外光谱更复杂,因为 H_2O 会在同一区域发生吸收,所以因此这种测量通常在 D_2O 中进行。除此之外,它们比紫外谱测量需要更多的蛋白质,溶液中蛋白质浓度要求更高。

第8讲

统计物理原理。温度与能量和熵变化的关系。不同能量状态的概率(玻尔兹曼-吉布斯分布)。配分函数及其与自由能的关系。构象转变。一级相变(“全或无”转变)和非相变的概念。构象转变过程克服自由能垒的内能。绝对反应速度理论的概念。并行过程和串行过程。扩散过程的特征速度。

现在可以开始转到讲述二级结构的形成与破裂。

然而,首先我想“大体上”谈一谈统计物理和热力学的基础知识,因为没有这些就很难讲述二级结构的稳定性、蛋白质的稳定性、多肽和蛋白质中的协同转变,以及这些转变的动力学。

热力学让我们了解在由极大量粒子组成的系统中可能存在的协同转变的类型。统计物理从所观察的粒子的性质和它们之间的相互作用出发,告诉我们在所观察的系统中什么时候会发生什么样的转变,并能描述这些转变的细节。

首先我们来看统计物理和热力学的基本概念:熵、温度、自由能和配分函数。

自由度数目巨大的系统(即由巨大数目的分子组成,或甚至仅由单个巨大的柔性分子组成)要使用统计物理来描述。“统计”是因为这样的大系统的构型(configuration)数量是极其巨大的。仅举一个例子:如果一条链的 N 个链节中的每一个都可以处于两种可能构型(如“螺旋”和“伸直”),则整个 N 节链有 2^N 个可能的构型。那么“普通的”含有 100 个链节的蛋白质链具有至少 2^{100},或大约 1 000 000 000 000 000 000 000 000 000 000 个构型。这是极多的。假设查看每个构型花费 1 纳秒(10^{-9}s),则把它们全部浏览一遍需要 3×10^{13}年,这比我们的宇宙存在的时间还要长 1000 倍……而在实验中的试管里有上百万条这样的链,还不算溶剂!如果我们想列出这些链的所有构型的清单,那我们就会耗费一辈子的时间。但是我们,当然,只对更简单和理性的东西感兴趣。例如,链的平均(即**统计平均**)螺旋度,以及这种螺旋度如何随温度而变化。而统计平均(即忽略所有不重要的细节)的卓越性质在于对于事件进行基本的简化。

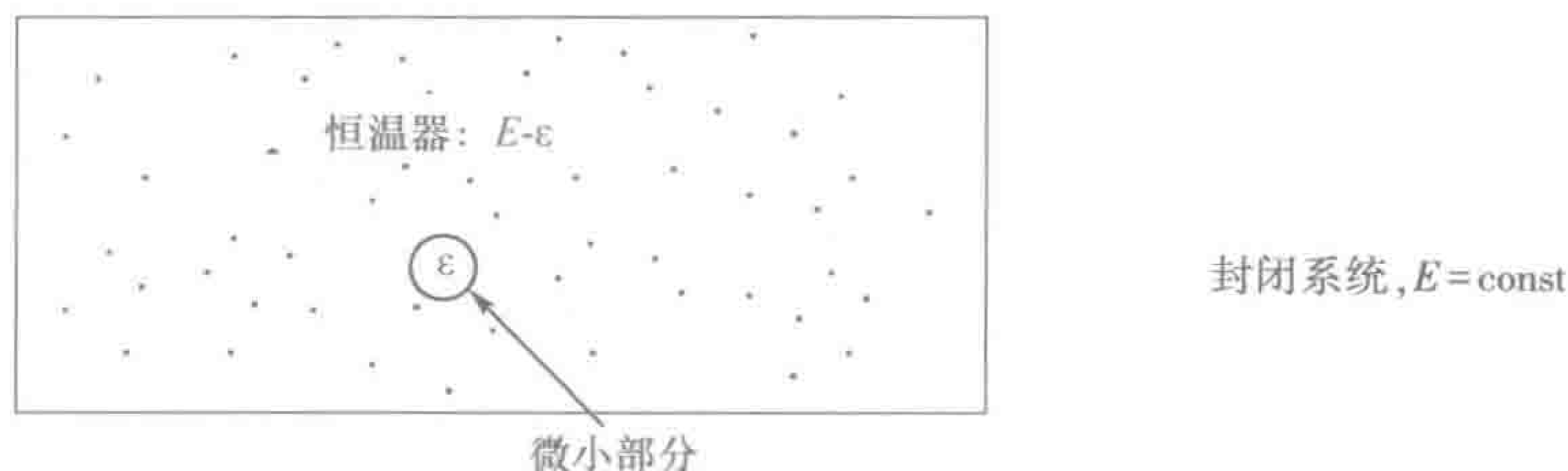
在统计平均中起主要作用的是**熵**。它说明所研究的系统有多少个构型(或者,如通常所说的,它有多少个**微观状态**)对应着我们能观察到的**宏观**(按照观察时间和同时被观察的分子数进行平均)**状态**。我们前面已经讲过一个类似的例子,但是那里我们讲到的实际上只是一种特例:“微观状态”的数目由分子所占据的体积所决定,而它的熵与该体积(而“处于给定的体积中”,如在这个房间里,即分子的“宏

观状态”)的对数成正比。

此处是合适的时候来回答一个很自然的问题:为什么物理学家更喜欢讨论微观状态数的**对数**,而不是状态数本身? 原因在于,如果讨论许多分立系统(如独立的分子),则我们可以看到,它们的体积、能量、自由度数这时可以相加,而可能状态数需要相乘。这种差别并不方便,它会产生繁琐的计算。但是可能状态数进行相乘时,其对数却可以相加(即 $\ln(AB)=\ln A+\ln B$),就像能量或体积一样。这种同类型性在解题时很方便,而且在加法操作时也让我们可以使用强有力的微分计算。

■ 现在来看**温度**。它与熵密切相关:**没有极大数目的状态,就没有熵,也不会有温度**。为了阐明它们之间的联系,让我们来考察一个封闭(即与外界没有能量交换)系统。设它的总能量为 E,并设该系统处于平衡,即它的所有能量为 E 的微观状态的概率是相等的(具有其他能量的微观状态的概率为零)。

在这个系统中划分出“被观察的微小部分”(如气体中的单个分子,或者一个大分子和其周围液体)。系统的所有剩余部分现在可以看做是恒温器,我们的“微小部分”就浸没于其中。



封闭系统,$E=\text{const}$

把系统的所有总能量为 E 的微观状态按照我们的“微小部分”的状态不同而进行分类。在每一类中系统微观状态数越多,则观察到该类的概率越大,即我们的“微小部分”的该状态的观察概率越大。

假设我们的“微小部分”固定在某个微观状态(如假设所观察的气体中的分子具有特定的空间位置和特定的速度)。设它在该状态下具有能量 ε。因为整个系统在整体上是封闭的,则**整个系统的总能量守恒**(能量守恒定律!),恒温器的能量等于 $E-\varepsilon$。设该能量对应的恒温器的可能微观状态数为 $M_{\text{therm}}(E-\varepsilon)$。则我们的“微小部分”的该状态的观察概率就很简单地正比于 $M_{\text{therm}}(E-\varepsilon)$。[备注:在这里我们隐含地认为,我们的系统的具体微观并不影响恒温器的微观状态。严格地讲,这并不是完全正确的(更准确点:这只是对于由稀薄的理想气体构成的恒温器才是正确的),但考虑到所有现象处于“我们的系统”的范围之内,恒温器可能会把后文的所有叙述搞乱,为了兼顾严格性和直观性,在我们的开始时可以假定我们“所观察的部分”封闭在固定的、恒温器分子不可穿透的体积内。在以后如果需要的话(对我们来说,是不需要的……),可以研究“所观察的部分”与恒温箱之间的相互作用。]

如果恒温箱的可能微观状态数等于 $M_{\text{therm}}(E-\varepsilon)$，则它的熵，**根据定义**，正比于这个数的对数：

$$S_{\text{therm}}(E-\varepsilon)=k\ln[M_{\text{therm}}(E-\varepsilon)] \tag{8.1}$$

此处的系数 k 是为了让熵像通常那样以 cal/K 为单位；后面我们会看到，它就相当于玻尔兹曼常量。

“微小部分”的能量 ε 也应该相当小。因此我们可以对 $S_{\text{therm}}(E-\varepsilon)$ 按参数 ε 应用常微分泰勒展开[当然，你们应该记得，在 dx 很小时，$f(x_0+dx)=f(x_0)+\frac{1}{1!}\left(\frac{df}{dx}\right)\Big|_{x_0}dx+\frac{1}{2!}\left(\frac{d^2f}{dx^2}\right)\Big|_{x_0}(dx)^2+\cdots=f(x_0)+\left(\frac{df}{dx}\right)\Big|_{x_0}dx$，其中 $\left(\frac{df}{dx}\right)\Big|_{x_0}$ 指的是微商 $\frac{df}{dx}$ 在 x_0 点处的值]。于是有：

$$S_{\text{therm}}(E-\varepsilon)=S_{\text{therm}}(E)-\left(\frac{dS_{\text{therm}}}{dE}\right)\Big|_{E}\cdot\varepsilon \tag{8.2}$$

[备注。因为 S 和 E 都正比于粒子数，所以 $\frac{dS_{\text{therm}}}{dE}$ 不依赖于恒温箱中的粒子数，同时 $\frac{d^2S_{\text{therm}}}{dE^2}$ **反比**于粒子数，即在极大的恒温箱中 $\frac{d^2S_{\text{therm}}}{dE^2}\to 0$；这就允许我们在式(8.2)中忽略 ε^2 项、ε^3 项和更高次项。]

这意味着，恒温箱的可能微观状态数与我们的“微小部分”的能量 ε 的依赖关系为

$$M(E-\varepsilon)=\exp\left[\frac{S_{\text{therm}}(E-\varepsilon)}{k}\right]=\exp\left[\frac{S_{\text{therm}}(E)}{k}\right]\times\exp\left[-\varepsilon\cdot\frac{dS_{\text{therm}}}{dE}\Big|_{E}\cdot\frac{1}{k}\right] \tag{8.3}$$

这里公倍数 $\exp\left[\frac{S_{\text{therm}}(E)}{k}\right]=M(E)$ 和 $\frac{dS_{\text{therm}}}{dE}\Big|_{dE}$ 值都与 ε 无关，与我们的“微小部分”的具体微观状态也完全无关。

由于微观状态数应该随能量增加而增加(能量越高，可分类方式的数目越多)，因此式(8.3)的意义很简单：我们的“微小部分”从恒温器中划出的能量越多，则留在恒温器的能量越小，而且可分类方式的数目越少。而且恒温器的可能微观状态数(其能量的分类方法数目)的减少与我们的“微小部分”中所包含的能量呈**指数**依赖关系。

结论：我们的“微小部分”(分子，等等)的特定微观状态的观测概率正比于 $\exp\left(-\varepsilon\cdot\frac{dS_{\text{therm}}}{dE}\Big|_{E}\cdot\frac{1}{k}\right)$，其中 ε 为该“微小部分”的能量，而数值 $\frac{dS_{\text{therm}}}{dE}\Big|_{E}\cdot\frac{1}{k}$ 与“微小部分”无关，仅与其周围环境的平均性质有关。

但是，根据玻尔兹曼公式，分子处于能量为 ε 的特定状态的概率正比于

$\exp\left(-\frac{\varepsilon}{k_B T}\right)$（其中 T 为温度，k_B 为玻尔兹曼常量）！比较 $\exp\left(-\varepsilon \cdot \left.\frac{dS_{therm}}{dE}\right|_E \cdot \frac{1}{k}\right)$ 和 $\exp\left(-\frac{\varepsilon}{k_B T}\right)$，可得

$$\left.\frac{dS_{therm}}{dE}\right|_E = \frac{1}{T} \tag{8.4}$$

而如果按通常方式，能量以焦耳（或卡）为单位，温度以 K 为单位，熵以 J/K（或 cal/K）为单位，则式（8.1）和式（8.3）中的 k 正是玻尔兹曼常量 k_B。

式（8.1）和式（8.4）是统计物理和热力学中的基本公式：它们定义温度为系统的熵 S（或微观状态数的对数）随温度的变化率的倒数。

特别地，它们还表明：$\ln[M(E+k_B T)] = \frac{S(E+k_B T)}{k_B} = \frac{S(E)+k_B T/T}{k_B} = \ln[M(E)]+1$，能量增加 $k_B T$ 时微观状态数增加 e=2.72 倍（接近 3 倍），这与系统大小、系统内部作用力等等都**无关**。

它们还表明，如果已知任意大的系统（“恒温器”）在不同能量时的微观状态数（或者说系统的“能量谱密度”），更准确些，如果已知系统能量谱密度与能量的对数关系，则我们可以求出该系统的每一能量所对应的温度。相应的图片显示于图 8-1 中。

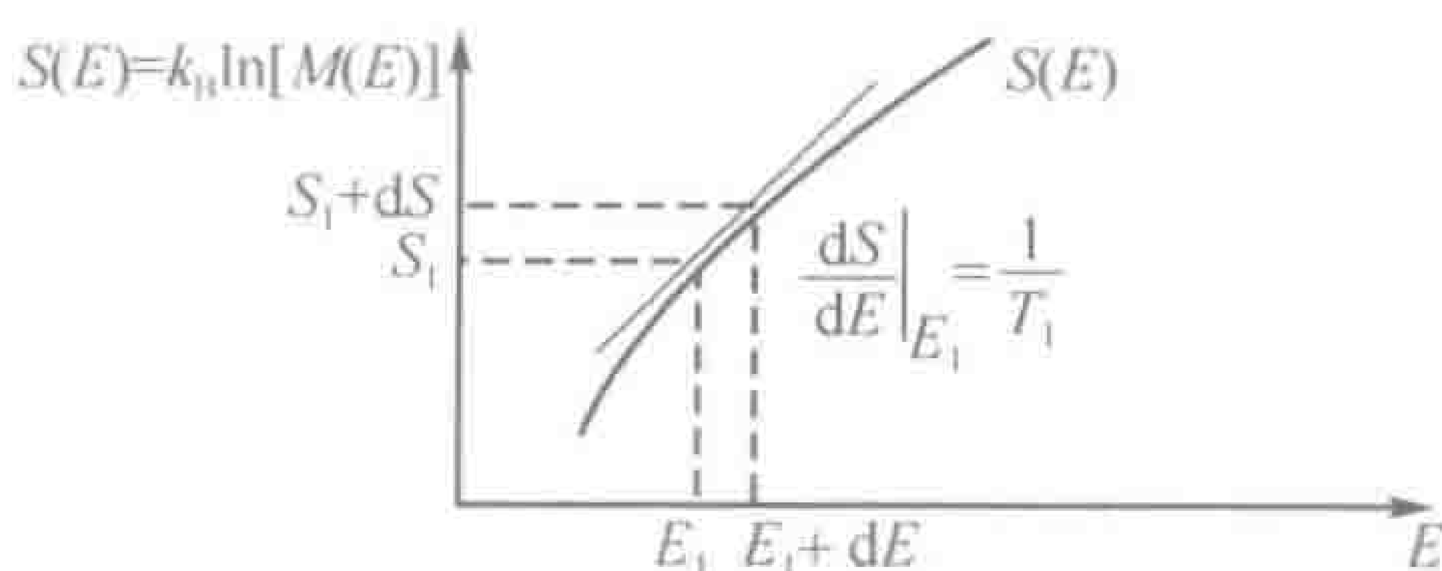


图 8-1　大系统的温度的确定。该系统的熵 S 与其能量 E 的关系曲线以粗线条显示。该曲线的斜率 dS/dE 确定给定能量 E 处所对应的温度 T。$M(E)$ 为给定能量 E 下的微观状态数，即系统状态的能量谱密度

所必需的是，微观状态应该非常多，因为只有当一个很小的（与 $k_B T$ 相比）能量间隔包含大量微观状态时，才能取对数 dS/dE。正因为如此，只有在研究极大的可能微观状态数目的学科中才出现温度。

■　让我们继续讨论封闭于固定的、恒温器分子无法穿透（但是通过器壁可以与温度等于 T 的恒温器交换能量）的体积中的系统。

式（8.1）和式（8.3）表明，在温度为 T 时我们的系统处于能量为 ε_i 的给定微观状态 i 的概率为

$$w_i(T) = \frac{\exp(-\varepsilon_i/k_B T)}{Z(T)} \tag{8.5}$$

其中

$$Z(T) = \sum_j \exp(-\varepsilon_j/k_B T) \tag{8.6}$$

为归一化因子，因为考虑到所有状态的概率之和 $\sum_j w_j$ 应该等于 1（此处和上面的

求和 $\sum_j$ 是对所研究的“微小系统”的所有可能微观状态 j 求和)。

数值 Z 被称为所研究系统的**配分函数**。若已知它,则可以按照式(8.5)计算在给定温度下该系统的每一个微观状态的概率,而且更进一步,在该温度下的系统平均能量:

$$E(T) = \sum_j w_j \varepsilon_j \tag{8.7}$$

以及它的平均熵

$$S(T) = k_B \sum_j w_j \ln(1/w_j) \tag{8.8}$$

注意,式(8.8)取遍系统的所有状态 j,按照它们(状态 j)的概率 w_j 对 $\ln(w_j)$ 取平均。这一公式是对我们已经熟悉的[公式(8.1)]计算熵的公式 $S=k_B \cdot \ln[M(E)]$ 在如下情形下的直接推广:系统的所有状态具有完全不同的概率——而不是由能量守恒定律推导出的仅有两个概率(对 $E_j=E$ 的所有 $M(E)$ 个状态有 $w_j=1/M(E)$,$E_j \neq E$ 时 $w_j=0$)。

课堂里的声音:我觉得,式(8.8)的含义应该更好地解释,应该证明,里面的 $S(T)/k_B$ 是平均系统状态数的对数。

讲课人:那么请忍耐一下,来看一些数学……

为了证明,来看 N(非常巨大的数目)个系统,其中的每一个处于状态"1"的概率是 w_1,处于状态"2"的概率是 w_2,……,处于状态"J"的概率是 w_J,这意味着,平均来说,N个系统中的 $n_1=w_1N$ 个处于状态"1",$n_2=w_2N$ 个处于状态"2",如此等等(而且 $\sum_j w_jN=N$)。有多少种方法把这 N 个系统分为 J 类,让 n_1 个处于状态"1",n_2 个处于状态"2",依此类推?

这个计算按如下进行。给系统分别编号为 1,2,……,N。现在从中挑出处于状态 1 的 n_1 个。有多少种不同的方式来选择?因为选取第 1 个时我们是从 N 个系统中进行选择,选取第 2 个时是从剩下的 $N-1$ 中进行选择,……,第 n_1 个是从剩余的 $N-(n_1-1)$ 中选择,因此,**按照这种选择顺序**,我们从 N 个系统中选取 n_1 个的方式有 $N(N-1)\cdots[N-(n_1-1)]$ 种。但是从系统中按什么顺序选择对我们并不重要:先取第 5 个,再取第 10 个,或者相反,这并不是**不同**的选择,它们的**结果**是**完全一样**的。而因为 n_1 个选出的系统可以按 $n_1(n_1-1)\cdots1$ 种不同操作顺序选取出来[开始可以选取所有 n_1 个中的1个,然后是从剩下的(n_1-1)个中选1个,依此类推],则从 N 个系统中选取 n_1 个总共有 $\{N(N-1)\cdots[N-(n_1-1)]\}/\{n_1(n_1-1)\cdots1\}$ 种**不同**的结果。把 $n(n-1)\cdots1$ 写作 $n!$(读做“n 阶乘”),可以看到,从 N 个中选取 n_1 个能给出 $N!/[(N-n_1)!\,n_1!]$ 种不同的结果。然后我们应该从剩余的 $N-n_1$ 个中选取 n_2 个。按上所述,在前一步从 N 个系统中选取 n_1 个的每一种方式下都可以给出 $(N-n_1)!/[(N-n_1-n_2)!\,n_2!]$ 种方式。因此,从 N 个系统中选取处于第 1 种状态的 n_1 个系统和处于第 2 种状态的 n_2 个系统可以有 $\{N!/[(N-n_1)!$

$n_1!]\}\{(N-n_1)!/[(N-n_1-n_2)!\ n_2!]\} = N!/[(N-n-n_2)!\ n_2!\ n_1!]$种不同的方式。继续这个推论链，可以看到，我们的 N 个系统可以有 $N!/[n_J!\cdots n_2!\ n_1!]$ 种分类方法，使其中的 n_1 个处于状态"1"，n_2 个处于状态"2"，……n_J 个处于状态"J"（提醒一下，$n_J = N-n_1-n_2-\cdots-n_{J-1}$）。倒数第二步需要努力的是：应用斯特林（Stirling）公式：$n! \approx (n/\mathrm{e})^n$。得到

$$N!/[n_J!\cdots n_2!\ n_1!] \approx (N/\mathrm{e})^{(nJ+\cdots+n1)}/[(n_J/\mathrm{e})^{nJ}\cdots(n_1/\mathrm{e})^{n1}]$$
$$= (N/n_J)^{nJ}\cdots(N/n_1)^{n1} = (1/w_J)^{NwJ}\cdots(1/w_1)^{Nw1} \quad (8.8a)$$

在最后一个等式我们应用了 $n_1 = w_1 N$ 之类的。因此，最后，让我们感兴趣的把 N 个系统按 J 种状态分类的方法数（让 n_1 个处于状态"1"，n_2 个处于状态"2"，依此类推）从式(8.8a)中可见为$[1/(w_J{}^{wJ}\cdots w_1{}^{w1})]^N$ 个。而由于这个数目（对于 N 个独立系统）就是单个系统的状态数的 N 次方，则单个系统的状态数目等于 $1/(w_J^{wJ}\cdots w_1{}^{w1})$，而它的对数（如所期待的那样）就是在式(8.8)中的 $S(T)/k_B$ 值。

得证！

配分函数在统计物理中具有重要的作用，因为 $Z(T)$ 值让我们能直接计算处于固定体积中的系统的自由能：

$$F(T) = E(T) - TS(T) = \sum_j \{w_j\varepsilon_j - T[-k_B \sum_j \ln(w_j)]\} = -k_B T \ln[Z(T)] \quad (8.9)$$

[在倒数第二等式我们使用了式(8.5)]。然后可以直接求得系统的熵：$S(T) = -\mathrm{d}F/\mathrm{d}T$[我们已经知道这个公式了，请读者从式(8.9)、式(8.6)、式(8.5)和式(8.8)给出的定义出发，自己推导一下]，再然后，可以得出它的能量 $E(T) \equiv F(T) + TS(T) = F(T) - T(\mathrm{d}F/\mathrm{d}T)$。

1）如果"微小系统"自身的自由度很大，则它具有自己的熵，从而具有自己的内部温度。"微小系统"的这个内部温度 T_{in} 等于恒温箱的温度 T。根据上述定义和等式，可如下推导：

$$T_{in} = \frac{\mathrm{d}E(T)}{\mathrm{d}S(T)} = \frac{\mathrm{d}\left[F(T) - T\frac{\mathrm{d}F}{\mathrm{d}T}\right]}{\mathrm{d}\left(-\frac{\mathrm{d}F}{\mathrm{d}T}\right)} = \frac{\left(\frac{\mathrm{d}F}{\mathrm{d}T} - T\frac{\mathrm{d}^2F}{\mathrm{d}T^2} - \frac{\mathrm{d}F}{\mathrm{d}T}\right)\mathrm{d}T}{-\frac{\mathrm{d}^2F}{\mathrm{d}T^2}\mathrm{d}T} = T$$

2）总能量等于内能和势能相加。前者仅取决于粒子的速度，而后者仅取决于它们的空间位置，而与速度无关。每个粒子的"微观状态"由它的空间位置和其速度所决定。在经典的（非量子的）力学中速度与坐标的配对是任意的[正如我们已知，海森堡量子关系 $\Delta v \Delta x \sim \hbar/m$，在室温下只是限制非常轻的粒子（电子）的速度与坐标配对关系]。这意味着，坐标与速度的概率分布可以"分离"，即 $w(\varepsilon_{内能}+\varepsilon_{坐标}) \sim \exp(-\varepsilon_{内能}/k_BT)\cdot\exp(-\varepsilon_{坐标}/k_BT)$。继续几步计算（请自己完成！），容易看到，自由能、能量和熵也可以"分离"为粒子的内能部分和坐标部分，即 $F = F^{内能} + F^{坐标}$，如此等等。重要的是，内能部分与系统构型**无关**（实际上，它们仅在热容方面提供了恒定的贡献），而且在研究旋转构象中可以把它们舍弃。因此在后文中我们将只讲述关于构型（或构象）部分的能量谱、能量及熵等。

3）上面我们按微观状态进行累加，但是在经典力学框架内我们可以按每个粒子的微观状态所对应的坐标和速度进行积分，也可以得到同样的结果。

4）平衡温度**只能是正的**。否则按照速度的概率积分，即 $\int \exp\left(-\frac{mv^2}{2kT}\right)\mathrm{d}v$，在速度很大时会趋向于无穷大，系统会“爆炸”。

因此在构象能量谱密度（同时，意味着系统的熵）随它的能量增加而降低的构象能量谱区域中，永远不可能观察到稳定状态：这些区域对应于 $T<0$［式(8.4)、式(8.1)和图 8-2(a)］

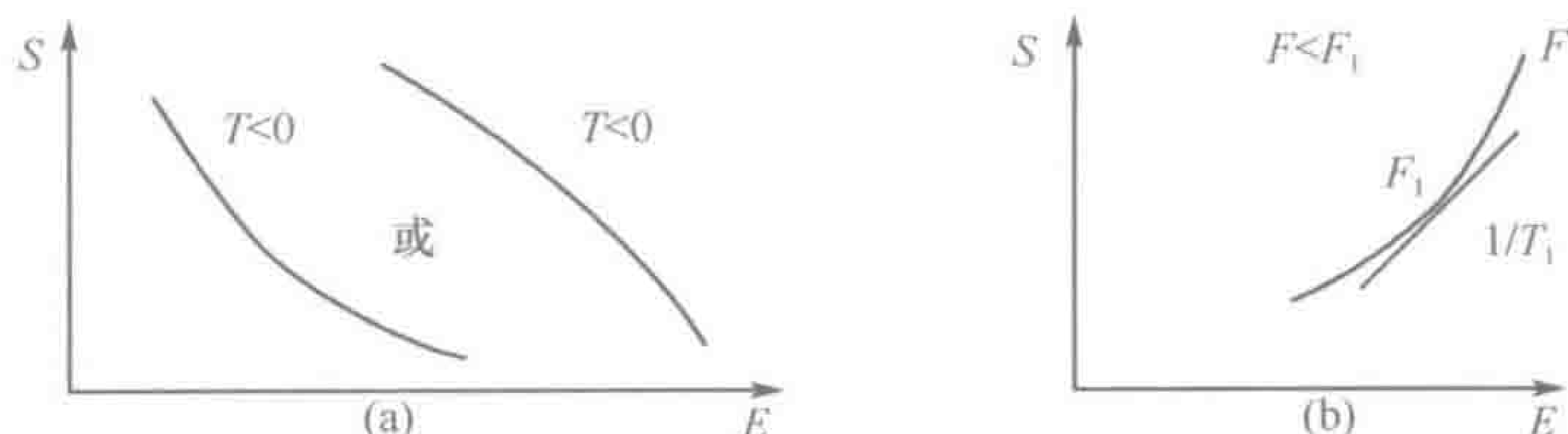


图 8-2　不可能对应系统的任何稳定状态的 $S(E)$ 图像局部。(a)熵 $S(E)$ 随能量 E 的增加而减小的区域：此处 $T=1/(\mathrm{d}S/\mathrm{d}E)<0$；(b) S 与 E 的关系曲线的“凹下”部分：在相切点($\mathrm{d}S/\mathrm{d}E=1/T_1$)自由能 $F=E-TS$ 不是小于，而是大于与之相邻的曲线 $S(E)$ 部分上的点，注意与图 8-3 对比

5）量 k_BT 的单位是“每粒子的能量”或“每摩尔($=6.02\times10^{23}$)粒子的能量”。如果人们从最开始的时候就把温度以能量单位来表示，则玻尔兹曼常数 k_B 根本就不需要。但是历史就是这样发生的，首先人们引入了温度单位“度”，后来才明白，“度”总是可以重新计算，转化为“每个粒子的能量”：对此只需把它(度)乘以某个常数——玻尔兹曼常量。相应地，k_B 的量纲是“每粒子每度的能量”。数值 k_B 取决于如何表示能量：或者是“每粒子的焦耳数”，或“每摩尔粒子的卡数”。根据度的能量值计量，$k_B=1.38\cdot10^{-23}$(焦耳/粒子)/度。不过 k_B 可以按另外的方法表示：不是按每个粒子计算，而是按每摩尔粒子计算：$1.38\cdot10^{-23}$(焦耳/粒子)/度$=1.38\cdot10^{-23}$(焦耳×$[6.02\cdot10^{23}]$/$[6.02\cdot10^{23}$个粒子$]$)/度$=[1.38\cdot10^{-23}\times6.02\cdot10^{23}]$(焦耳/摩尔粒子)/度$=8.31$(焦耳/摩尔)/度$=1.99$(卡/摩尔)/度。最后的数值，1.99(cal/mol)/K，传统是被称为“理想气体常数”R。

6）比熵(specificentropy)常常以熵单位“e. u.”来表示：e. u. $=$(cal/mol)/K$\approx k_B/2$。因此，1e. u. 的熵对应于每粒子 $e^{1/2}\approx1.65$ 个系统状态。

■　现在，我们终于可以讨论**构象转变**了。正是为了对它有足够的讨论，我们才回顾了统计物理和热力学的基础知识。

构象转变分为渐变的(“连续变化”)和急剧的(“相变”)。通过使用图 8-1 所示的能量状态密度图，我们来试着弄清楚它们的区别。因此首先需要学会如何确定图中的稳定态，或者我们的系统在给定的环境温度下的状态。

假设已经给定我们的系统(“大分子”)在每一个能量 E 下的状态数 $M(E)$ 和周围环境(“恒温器”)的温度 T_1。如果已知熵曲线 $S(E)=k_B\ln[M(E)]$ 和 T_1，如何在曲线上找到与系统的稳定态所对应的点？温度 T_1 表明了在曲线上我们所关

心的点的斜率 dS/dE：$dS/dE=1/T_1$，因为我们的大分子的温度与其环境的温度相同。

但是，通常来说，在曲线具有给定斜率 $1/T_1$ 的点会有多个（见下文，图 8-5）。哪一个才对应着稳定状态？来看图中 E_1 处的切线，在该处有$\left.\frac{dS}{dS}\right|_{E1}=\frac{1}{T_1}$。该切线的公式为 $S-S(E_1)=(E-E_1)/T_1$，或 $E-T_1S=E_1-T_1S(E_1)$。值 $F_1=E_1-T_1S(E_1)$ 就是在温度为 T_1 时的自由能。沿此切线，正如我们看到的，值 $(E-T_1S)$ 是不变的。切线左侧各点处的 $E-T_1S$ 值比切线上要低，而右边各处的值要高（图 8-3）。

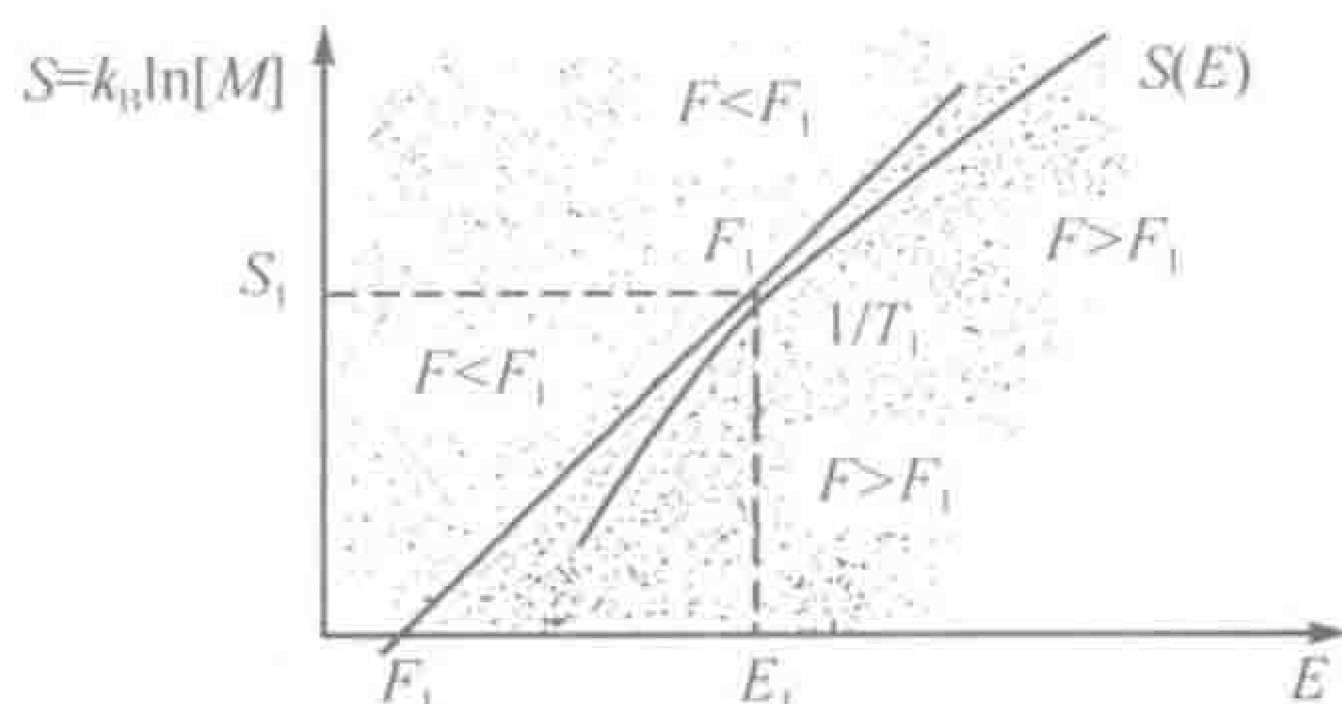


图 8-3　在图像上确定温度和自由能。熵 $S=k_B\ln(M)$ 与 T 的关系以曲线 $S(E)$ 表示。$M(E)$ 为系统状态的能量谱密度。曲线 $S(E)$ 的斜率 $dS/dE=1/T$ 能确定温度 T。物理上可能的温度 $T>0$ 保证了曲线 $S(E)$ 是上升的。F_1 为给定能量谱在温度为 T_1 时的自由能。数值 $F=E-T_1S$ 沿着切线为常数且等于 F_1。该切线的左边（上方）$F=E-T_1S<F_1$，右边（下方）$F=E-T_1S>F_1$。由于此处的曲线 $S(E)$ 是凸起的，即处于切线的右下方，则切点对应于温度为 T_1 时的自由能最小值。更详细的解释见正文

特别是，后者意味着 S 与 E 的关系曲线的"凹下"部分[图 8-2(b)]绝对不可能对应着稳定状态：在该处切点的 F 不是小于，而是大于 $S(E)$ 曲线上相邻区段的值。这意味着，如果从切点沿凹曲线 $S(E)$ 移动，系统会降低其自由能。

如果曲线 $S(E)$ 各处都是凸起的，则它的斜率总是随能量 E 的增大而减小。那么每一个斜率 $1/T$ 仅对于曲线上的一个点（图 8-1，图 8-4），则该点对应着给定温度 T 下的稳定态。随着温度的变化，该点逐渐移动，而系统也逐渐地改变自身的热力学状态，即它的熵和能量（图 8-4）。

如果曲线 $S(T)$ 的斜率随着能量 E 的增大时而减小，时而增大（图 8-5），则以相同的斜率与该曲线相切的可能有好几处，最左边的切线（其 F 比较小）反映了稳定状态。在这种情况下在某个温度 T^* 以下，在能量较低区域的某条切线会是"最佳"（对应着最小的自由能 F 值），在 T^* 之后"最佳"切线将会出现在能量较高区域。

在温度 T^*，低能结构和高能结构会具有相等的自由能和相等的存在概率。这意味着，如果在转变温度 T^* 存在大量相同的系统，我们将会观察到其中一半处于低能态，另一半处于高能态（图 8-5 右）。换句话说，每个系统有一半时间处于高能

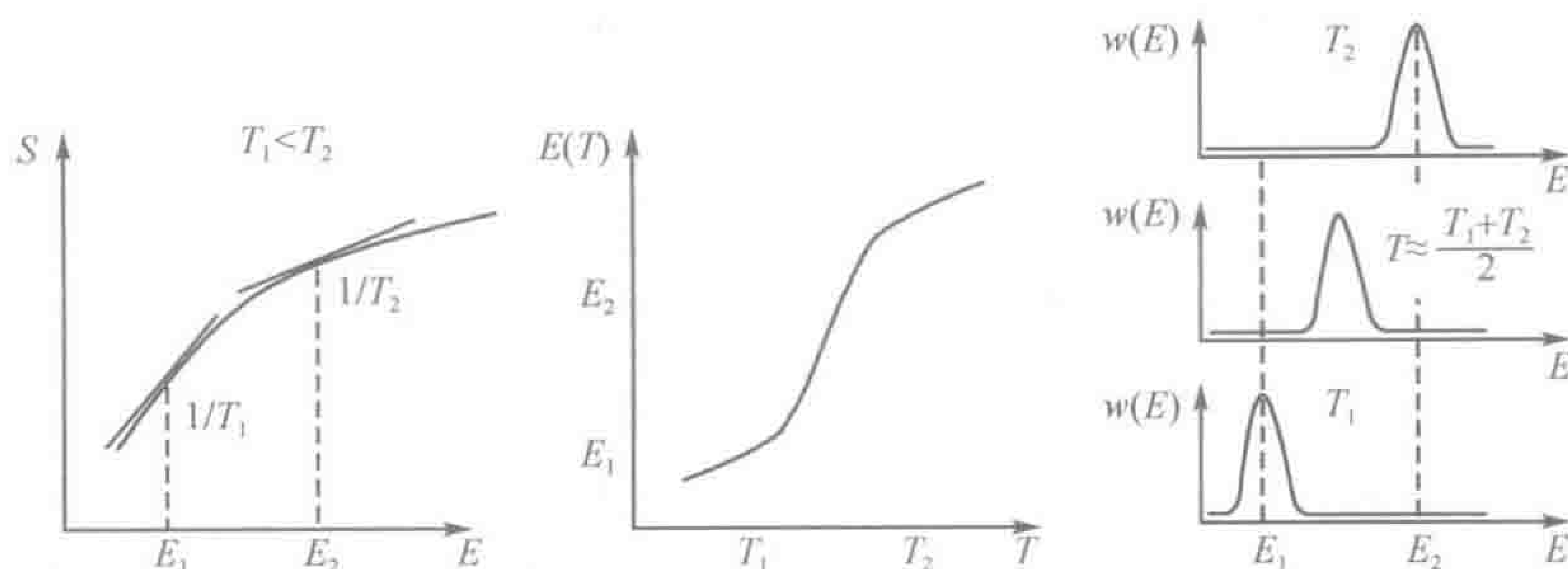


图 8-4　系统状态随温度变化而发生的逐渐变化。为此曲线 $S(E)$ 必须是凸起的。如果已知关系曲线 $S(E)$（左图），可以得到关系曲线 $T(E)$，进一步得到 $E(T)$，后者显示于中图。右图：$w(E)$ 为分子的观察状态按能量的分布，即在给定温度 T 时具有能量 E 的概率；概率 $w(E)$ 正比于 $\exp\left(-\frac{E-TS(E)}{k_B T}\right)$

态，另一半时间处于低能态。能量全然不同的两种可能状态的这种“共存”将会占据在 T^* 附近某个非常狭窄的（尤其是，对于巨大的“宏观”系统）的温度范围，这个范围我们很快就要进行估算。

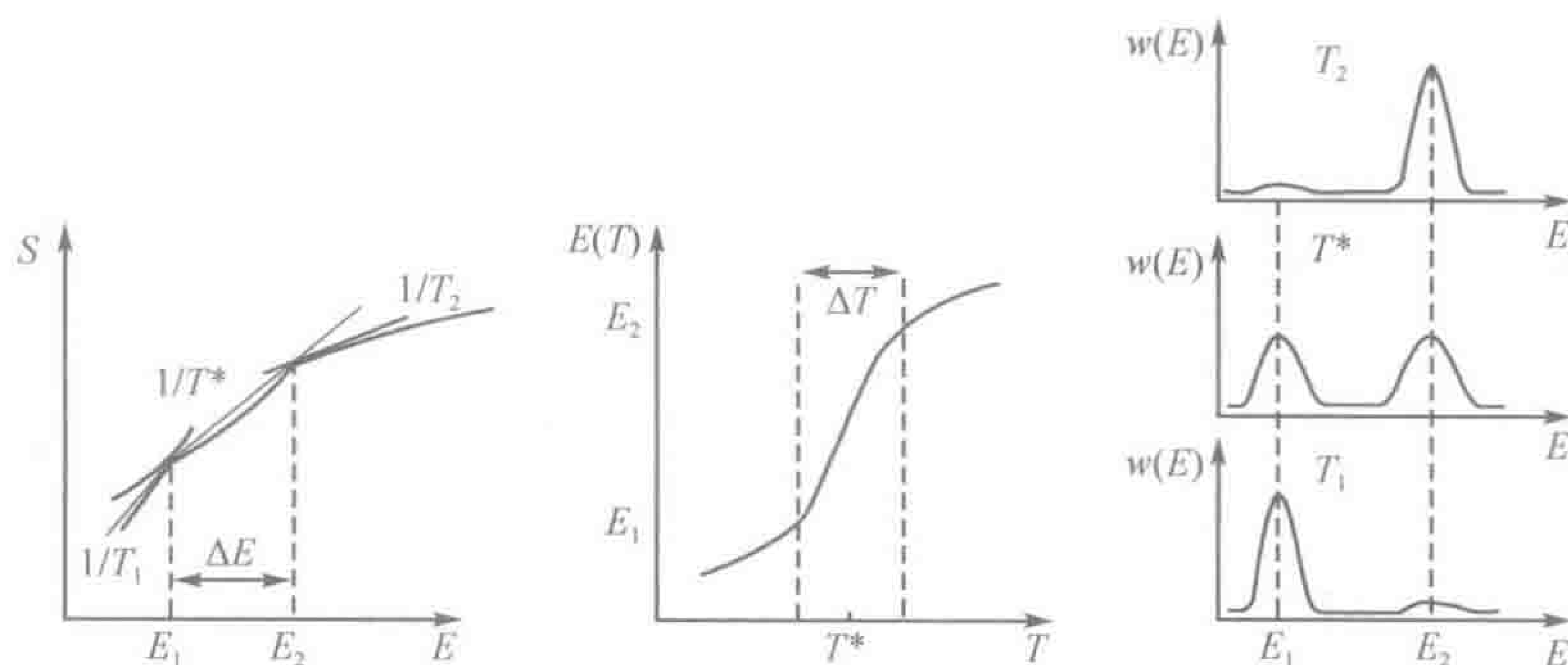


图 8-5　“全或无”型相变（在宏观物体中被称为“一级相变”）的特征是系统状态随温度改变而急剧变化。为此曲线 $S(E)$ 必须具有凹下的部分：在转变温度 T^* 下，该部分内的自由能比它的两侧要高[图 8-2(b)]。转变发生在对应着低能量状态与高能量状态共存（即概率大致相等）的狭窄的温度范围（ΔT）。切线对应着转变中点温度 T^* 以及 $T_1<T^*$ 和 $T_2>T^*$。注意，“全或无”型转变（即一级相变）的温度 T^* 恰好与能量的急剧变化的中点和分子按能量分布函数 $w(E)$ 的分裂极大处相对应

然而，最重要的是，具有“中等”能量的状态**不能**成为稳定状态：由于曲线 $S(E)$ 是弯曲的，图中对应于这些“中间”状态的点处于与温度 T^* 相对应的切线的**右边**，即在温度为 T^* 时这些“中间”状态的自由能比两个切点对应的结构要高，出现概率极小。于是可以说，这两个稳定状态被“**自由能垒**”所分隔。

这就是“**全或无**”转变的条件。

在微观系统(尤其是蛋白质球体)中这种转变可以通过温度 T^* 下在自由能垒上的跃变而发生;对此我们将在本讲的结尾处讨论。

在宏观系统中(如在装有冰水的试管中)能垒(在温度 T^* 下)太高,以致跃过它需要花费的时间是无穷大。因此在宏观系统中观察到有**滞后**:系统状态保持不变直到与温度 T^* 相比略微过冷(凝固时)或过热(熔化时),然后转变通过在系统产生瞬时的(即不稳定的)相分离(如液态和固态)而发生。在宏观系统中的这种转变被称为**一级相变**。

■ 现在我想估算高能结构与低能结构共存的温度区间的数值。在转变中点,在温度为 T^* 时,低能相的自由能为 $F_1(T^*)=E_1-T^*S_1$,等于高能相的自由能 $F_2(T^*)=E_2-T^*S_2$,即在转变中间,有

$$(E_2-E_1)=T^*(S_2-S_1) \tag{8.10}$$

在相对于 T^* 有一个微小的温度偏移(δT)时,相的自由能会发生改变,因为两相的自由能之差是 $\delta F=F_1(T^*+\delta T)-F_2(T^*+\delta T)=(dF_1/dT)\delta T-(dF_2/dT)\delta T=(-S_1\delta T)-(-S_2\delta T)=\delta T(S_2-S_1)$。两相共存(即它们概率大致相等,或者说,它们的概率比在有利于第一个相的 10∶1 到有利于第二相的 1∶10 之间变化时)的范围是:$\exp(-\delta F/k_B T^*)$ 在 10 至 1/10 之间,即 $\delta F/k_B T^*$ 处于 $\ln(1/10)\approx-2$ 至 $\ln(10)\approx+2$ 之间。在此区间 $|\delta T|<2k_B T^*/(S_2-S_1)$。意味着,相共存的温度间隔为

$$\Delta T \approx \frac{2k_B T^*}{S_2-S_1}-\frac{-2k_B T^*}{S_2-S_1}=\frac{4k_B T^*}{S_2-S_1}=\frac{2k_B(T^*)^2}{E_2-E_1} \tag{8.11}$$

来看一个有益的数值例子。如果 $T^*\sim300$°K(即 $k_B T^*\sim0.6$kcal/mol),而 E_2-E_1 取蛋白质熔融的典型值,约为 50kcal/mol[即 $50/(6\times10^{23})\sim10^{-22}$kcal 每个蛋白质],则 ΔT 约为 10 度。即熔融的和固态的蛋白质分子在转变中点周围的几度范围内共存。如果一个系统的 E_2-E_1 约为 50kcal(相当于瓶子大小的冰块融化时的能量),则相共存的区间 ΔT 约为 10^{-23}度。

换句话说,小系统中的"全或无"型相变和(在更大程度上)宏观系统中的一级相变,其特征是在狭窄的温度范围内的能量急剧跃变。在宏观系统中这个范围实际上是无限狭窄的,而对大分子它约为几度,即与"观测区间"(通常处在 0~100℃)相比也是很狭窄的。而在较小的低聚肽上没有剧烈的跃变:能量变化区间可能包含被研究温度的整个"实验窗"。

■ 几句关于**二级相变**的话。如果一级相变的特征是系统的**能量**(及其熵、体积和密度)**跃变**,那么二级相变的特征只是关系曲线 $E(T)$ 的**陡变**,即能量随温度而改变的速度的跃变,或者说,**比热容跃变**。

要强调的是,二级相变发生在曲线 $E(T)$ 的**陡变点**(而对于蛋白质类型的小系统,是在急剧拐弯的狭窄区域)(图 8-6),而绝对不是在紧随其后的大致为"S 形"的能量-温度关系曲线的中

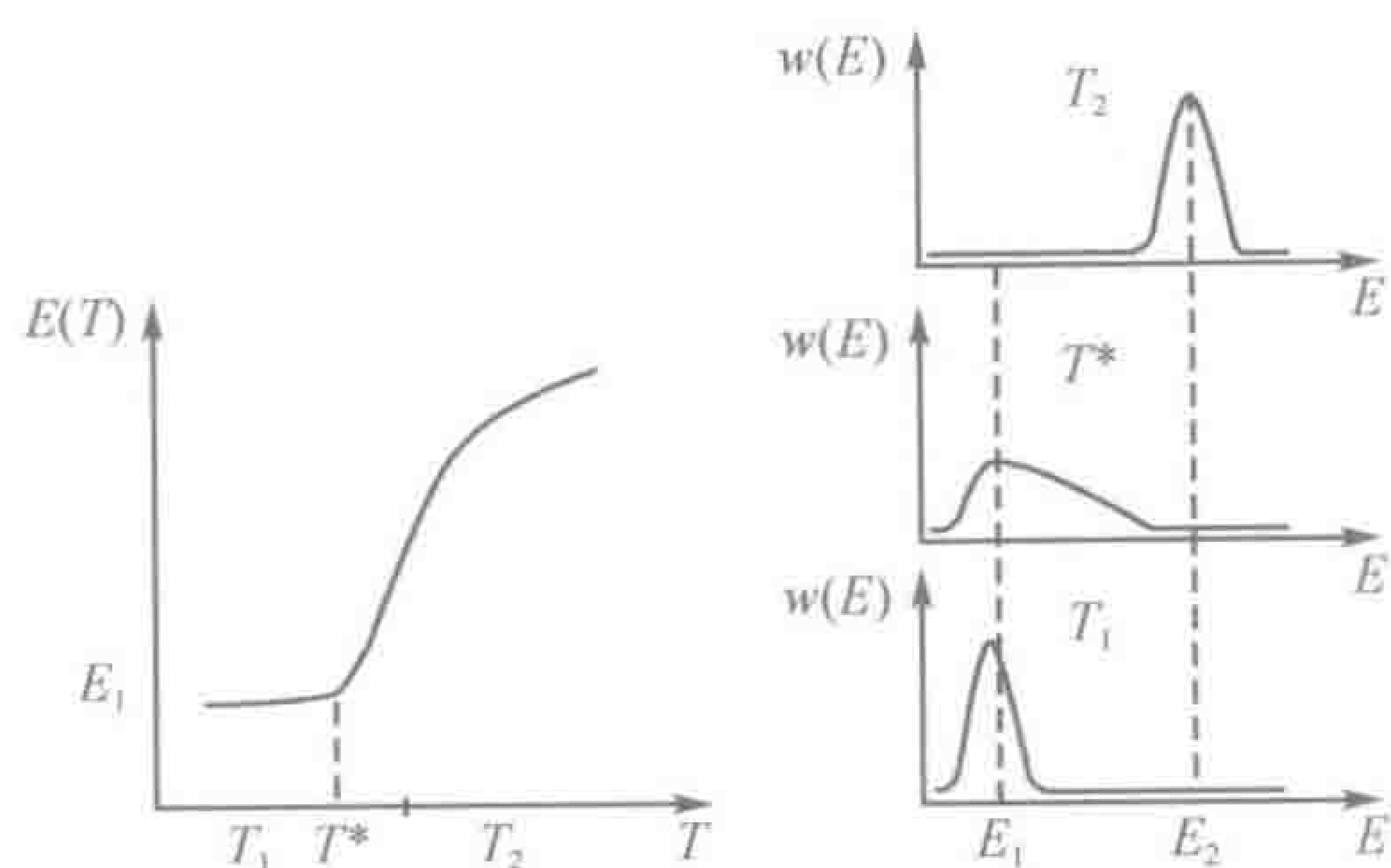


图 8-6 **二级相变**中的特征能量关系曲线 $E(T)$ 和观测状态的概率按能量的分布函数 $w(E)$。T^* 为该相变的温度。在该温度下分子按能量的分布函数 $w(E)$ 急剧地加宽，能量**开始**(或者**结束**，这取决于从哪边观察)急剧地变化。注意，二级相变的温度 T^* 恰好与能量急剧变化的**开始处**(而非中部)相重合

部。这种相变出现在，例如，当系统停留在特定的状态，直到(或从该温度开始)某个特定的温度(在系统加热时吸收的所有能量仅增强它的热振动)，但是在这个临界温度之后，部分热量被额外地用于系统的重建(例如在低于居里温度的铁磁材料中，被用于破坏其自发的磁化状态)。

在图 8-4～图 8-6 中我特地展示的是系统能量开始随温度缓慢变化，然后急剧变化，再后来又缓慢变化的情形。正如大家看到的，$E(T)$ 的这种"S 形"形状可以是一级相变，也可以是在完全没有任何相变时系统的逐渐变化，也可以是由二级相变触发的逐渐变化。

我请大家对此特别注意，因为在非物理学家中存在着两个不知从哪里来的错误想法："转变"总是出现在任何 S 形改变的中间；如果这种"转变"不是一级的，就是二级的。不是这样的！

注意：虽然在图 8-4～图 8-6 中的曲线 $E(T)$ 看上去，大体上是相同的，但按能量的分布曲线 $w(E)$ 在一级相变(即"全或无"型)的外观与其他转变**完全不同**。"全或无"型转变(而且只有它们)的特征是在 $w(E)$ 上存在有两个峰，即两相的**共存**。因此为了说明在两个极端状态之间如何(跃变型还是渐变型)发生转变，只看到能量或其他某种观察参数在狭窄的温度区间内急剧变化**是不够的**。这里需要额外的测量，对此我们将会在后面的某讲中讨论。

■ 在结尾我们来谈一谈构象转换的动力学。更准确一点，是关于为什么某些转换进行得非常缓慢。这里的"缓慢"是什么意思？假设你们已经知道过程中的一个基本步骤的时间。例如，一个残基在 1ns(纳秒)内折叠进入到二级结构。你们也知道，在链中有 1000 个残基。折叠过程不是进行 1000ns，而是 1000s。这就是"缓慢"：过程进行的时间的数量级比已知基本步骤的时间和步骤数量后所预期的要慢。需要理解的是，为什么会这样……

过程的"缓慢性"有时与高黏度物质的扩散缓慢性相关。但是更常见的是(强调一下，并非总是，但常常)，过程的缓慢性与需要克服很高的**自由能垒**有关。这对"**全或无**"型转变更为典型，其中的自由能垒把两相分隔开了(图 8-5)；其能垒的值总是很大。这种能垒与化学应中的活化能非常相像，只是在这种情形下它既具有能量成分，也有熵的成分。我来提醒一下，如何按照经典的过渡态理论估算这种

"能垒克服"反应的速度。

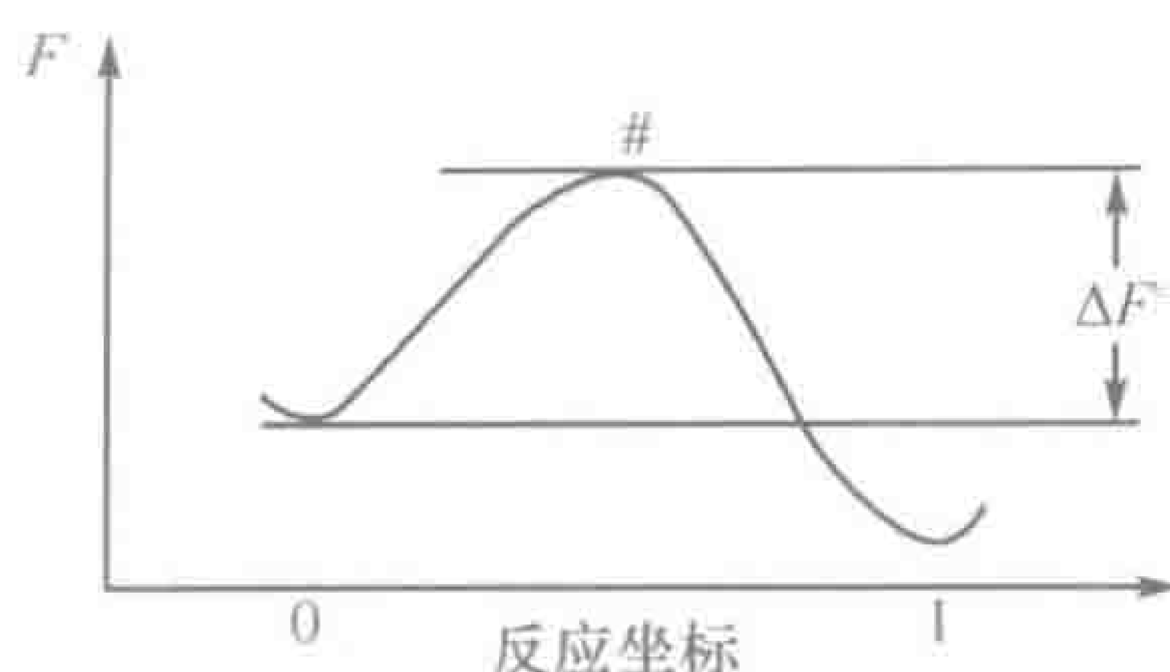


图 8-7　在从状态"0"转变到状态"1"中要克服自由能垒("活化能")#。$\Delta F^{\#}$为能垒的(过渡态的)自由能

首先来看系统从状态 0 转变到状态 1 的最简单过程。设在"0→1"过程的途径中有一个"能垒"#(图 8-7),在这一途径上,在它的"旁边"都没有任何"势阱"(图 8-8:状态 X,它比初态"0"更稳定,但不如末态"1"那么稳定;势阱的存在使过程更复杂,但是并不改变它的本质)。

于是,过程的速度取决于:①能垒态("过渡态")的密度;②从能垒态转变到"能垒后"状态的速率。

若 $\Delta F^{\#}$为从初态的自由能算起的能垒自由能,且 $\Delta F^{\#}>>k_B T$,而且如果没有"势阱"(图 8-8),若在初态有 n 个"粒子"(同义词:分子,系统等),则任一时刻,有(涨落的结果)$n^{\#} \sim n^{e}xp(-\Delta F^{\#}/k_B T)$ 个粒子处在能垒上。设在能垒上的每个粒子在时间 τ("基元反应步骤的时间")内跃过能垒。于是在约为 τ 的时间间隔内所有在能垒上的 $n^{\#}$个粒子都跃过能垒。则所有 n 个粒子翻过能垒需要 $n/n^{\#}$步,即时间约为

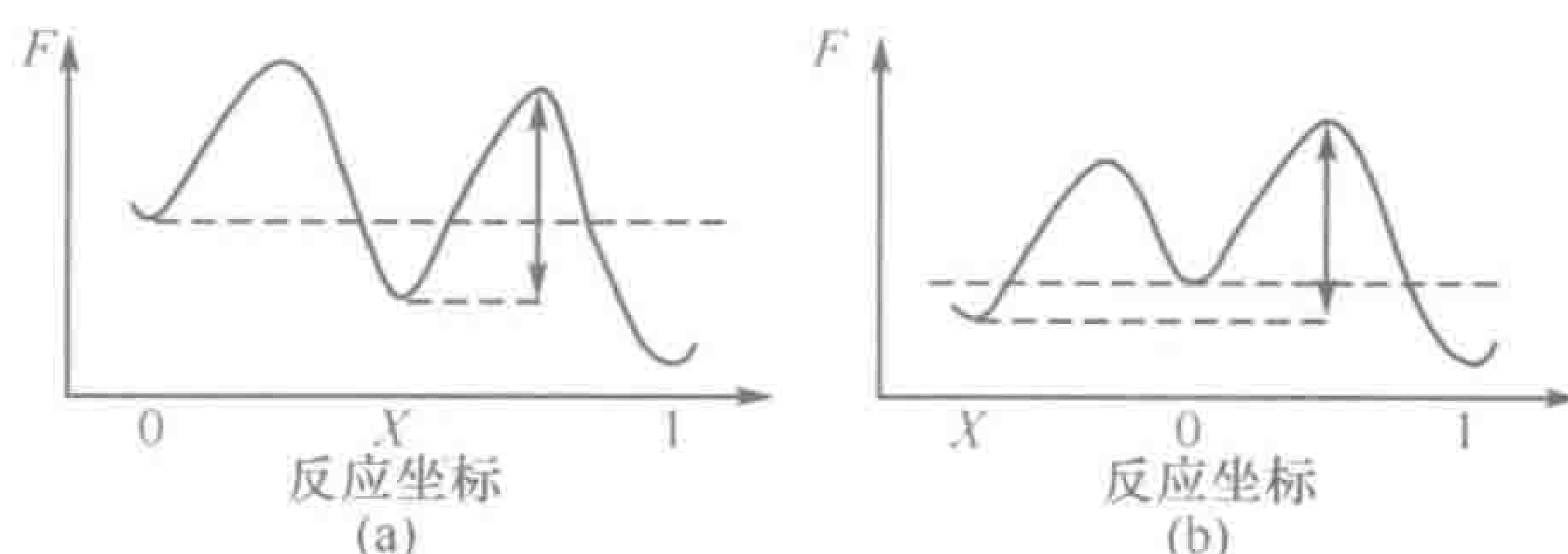


图 8-8　"势阱":它可以是路径 0→1 的中间体"X"(a),也可能是另外的(对反应 0→1 而言)状态"X"(b)。对动力学"势阱"来说重要的是,它比初态"0"更稳定,但没有末态"1"那么稳定,而且在"X"与"1"之间的自由能垒(图中箭头所示)比"X"与"0"之间的更高

$$t_{0\to1} \sim \tau(n/n^{\#}) \sim \tau\exp(+\Delta F^{\#}/k_B T) \tag{8.12}$$

注意,值 $k_{0\to1} \equiv 1/t_{0\to1}$,被称为从 0 到 1 的转变速度。

如果存在"势阱"X,可用同样的方法估算从 0 转变到 X 和从 X 转变到 1 的时间。而且从 0 转变到 1 的时间等于 $0\to X$ 和 $X\to1$ 的转变时间相加。

补充说明。

(1) 如果转变可以通过几条**并行的**途径进行,则并行转变的**速度**可以**相加**:

$$k_{1+2+}\cdots = k'_{0\to1} + k''_{0\to2} + \cdots \tag{8.13}$$

这里 $k'_{0\to1} \equiv 1/t'_{0\to1} = (1/\tau)\exp(-\Delta F^{\#1}/k_B T)$,是沿着并行路径中的第一条途径的转变速率,$k''_{0\to2}$是沿第二条路径的,如此等等。如果并行路径并不是太多,则过程时间由其中最快的(即能垒最低的)途径所决定。

(2)如果转变过程中需要**连串地**克服几个能垒,则需要把克服各个能垒的**时间累加**起来。这个结论在图 8-8(a)所显示的过程中很容易看出来,在该过程中稳



图 8-9　在并行过程(a)和串行过程(b)中的自由能垒

定中间体 X 首先累积起来,然后从它形成末态 1。但是对图 8-9(b)所示的过程来说,这并不容易看出,在该过程中所有的中间体都**不**稳定。除此之外,必须明确,在这种情形下的"克服单个能垒的时间"必须理解为从能垒之前的**最深**的自由能最小值升高到能垒的时间,而不是从直接相连的前一自由能最小值开始[因此,在图 8-9(b)所示的情形中,能垒#2 的高度应该从状态 0 开始算起,而不是从状态 1 开始]。

为了证明这一结论(我只给出思想而不拿计算折磨大家了),请看如下过程:

$$\boxed{\begin{matrix}0\\ \text{开始}\end{matrix}} \underset{k_{1\to 0}}{\overset{k_{0\to 1}}{\rightleftarrows}} \boxed{1} \underset{k_{2\to 1}}{\overset{k_{1\to 2}}{\rightleftarrows}} \cdots\cdots \underset{k_{M\to M-1}}{\overset{k_{M-1\to M}}{\rightleftarrows}} \boxed{M} \underset{k_{M+1\to M}}{\overset{k_{M\to M-1}}{\rightleftarrows}} \boxed{\begin{matrix}M+1\\ \text{结束}\end{matrix}}$$

(其中 $k_{i\to i+1}$ 为从 i 到 $i+1$ 的转变速率,$k_{i+1\to i}$ 为 $i+1$ 到 i 的逆转变速率),该过程中所有中间态($1,2,\cdots,M$)的自由能比初态 0 和末态 $M+1$ 要高出 k_BT 的很多倍。

由于所有中间状态的自由能较高,在这些中间状态中只有很少量的分子(与初态和末态的数量相比)。因此在每个中间状态的分子数量的变化率与初态和末态的分子数量变化率相比,要小得多。换句话说,可以假设,沿整个反应路径的状态流通率为

$$-\mathrm{d}n_0/\mathrm{d}t \equiv k_{0\to 1}n_0 - k_{1\to 0}n_1 = k_{1\to 2}n_1 - k_{2\to 1}n_2 = \cdots = k_{M\to M+1}n_M - k_{M+1\to M}n_{M+1} \equiv \mathrm{d}n_{M+1}/\mathrm{d}t \tag{8.14}$$

(这一假设被称为"准静态**近似**";它在化学动力学中被广泛地使用)。求解这些等式得到公式:

$$t_{0\to\cdots\to M+1} = (1/k_{0\to 1}) + (1/k_{1\to 2})\exp(\Delta F_1/k_BT) + \cdots + (1/k_{M\to M+1})\exp(\Delta F_M/k_BT) \tag{8.15}$$

式中,$1/k_{i\to i+1}$ 为克服第 i 个能垒的时间(从直接相邻的前一个状态 i 开始),而 ΔF_i 为中间态 i 的自由能(从初始状态计算起)。还有,需要提及的是,在推导出这个公式时我使用了熟知的关系式

$$k_{i\to j}/k_{j\to i} = \exp[(F_i - F_j)/k_BT] \tag{8-16}$$

它是这样推导出来的:这些状态的平衡密度,n_i^0 和 n_j^0,应该满足动力学方程 $n_i^0 k_{i\to j} = n_j^0 k_{j\to i}$(在平衡时从一个状态到另一个状态的粒子流为零),还要满足热力学关系式 $n_i^0/n_j^0 = \exp[-(F_i - F_j)/k_BT]$(把粒子处于两种状态的平衡概率与其自由能联系起来)。

还有最后一个条件。因为按照式(8.12),$1/k_{i-1\to i} = \tau_i\exp[(\Delta F^{\#i} - \Delta F_{i-1})/k_BT]$,则 $(1/k_{i-1\to i})\exp(\Delta F_{i-1}/k_BT) = \tau_i{}^{e}xp(\Delta F^{\#i}/k_BT)$。这就是穿越单个能垒#$i$ 的转变时间,如果这个能垒是在过程 $0\to M+1$ 的路径上唯一的能垒的话。我们记作 $t_{0\to\#i\to}$。则有

$$t_{0\to\cdots\to M+1}=t_{0\to\#1\to}+t_{0\to\#2\to}+\cdots+t_{0\to\#M+1\to} \tag{8.17}$$

这就证明了：**串行反应**的进行时间等于克服各个能垒的**时间加和**，其中能垒的高度从前面的**最深**的自由能最小值算起。

顺便提一句，式(8.17)表明，如果能垒不是太多，则过程的时间可简单地由其中**最高的**能垒所决定。

■ 最后。正如我已经讲过，只要当我们看到过程进行得比扩散慢得多的时候，我们就猜想存在着自由能垒。而扩散的进行需要多少时间？为了让大家有一个定性的印象，我们对扩散过程来讲一点点。

但是首先有必要理解，由于环境中的摩擦，分子需要多长时间才会忘记它的流动方向，并开始“扩散”，即需要理解分子动能因摩擦而耗散需要多长时间。

可以证明，这会在几皮秒(10^{-12}s)内发生。

在本质上，分子在黏性介质中的惯性运动可用如下简单公式描述：

$$m\frac{\mathrm{d}v}{\mathrm{d}t}=F_{\text{frict}} \tag{8.18}$$

式中，$\mathrm{d}v/\mathrm{d}t$ 为粒子的加速度；m 为粒子的质量；F_{frict}为它与周围介质的摩擦力。质量可以按公式 $m=\rho V$ 估算，其中 ρ 为粒子密度，V 为体积。可以合理地认为，粒子的所有线性尺寸大致相同并等于 D，于是有 $V\approx D^3$。摩擦力 F_{frict}可以按斯托克斯公式估算：$F_{\text{frict}}=-3\pi D\eta v$，其中 v 为粒子运动速度，η 为介质黏度。则公式

$$m\frac{\mathrm{d}v}{\mathrm{d}t}=-3\pi D\eta v \tag{8.18a}$$

给出了特征时间：

$$t_{\text{kinet}}\approx m/(3\pi D\eta) \tag{8.19}$$

即在介质摩擦力作用下的运动阻尼时间。它很容易进行数值估算。因为 $m\approx\rho D^3$，而 $3\pi\approx10$，则 $t_{\text{kinet}}\approx0.1\rho D^2/\eta$。而且，因为我们所处理的所有分子的密度 $\rho\approx1\text{g/cm}^3$，而水的黏度 $\eta\approx0.01\text{g/(cm}\cdot\text{s)}$(见相关手册)，则

$$t_{\text{kinet}}\approx10^{-13}\text{s}(D/\text{nm})^2 \tag{8.19a}$$

式中，(D/nm)为以纳米为单位表示的分子直径。

这意味着，小分子($D\approx0.3$)的内能在~10^{-14}s 内耗散掉，较小的蛋白质($D\approx3$nm)约为 10^{-12} s，而大蛋白质($D\approx10$nm)约为 10^{-11}秒。因此：特征时间为皮秒(10^{-12}s)。在黏度比水更大的环境(如膜)的摩擦力作用下，内能耗散得更快。

当然，没有能量了我们的粒子并不会停止不动：介质分子从各个方向飞到我们的分子上，把自己的能量送给它。而它这时完全忘掉了自己最初的运动方向——在皮秒内……

备注：上述的 t_{kinet}对应于水环境。根据现有的数据，蛋白质分子的细胞环境的黏度大约高一个数量级。因此在那种环境中 t_{kinet}比纯水环境大约低一个数量级[式(8.19)]。

由热维持的任意分子的内能，$mv^2/2$，平均来说，正如我们知道的，其数值大致为 k_BT。分子在 t_{kinet}这段时间内记得自己的运动方向。这段时间如果作为一个步骤的时间，它行进的距离为 $\Delta l\sim vt_{\text{kinet}}$。然后它的运动方向发生改变，它在某个未定的方向上移动同样的距离 Δl，即它移动±

Δl。在这种情形下分子离开起始点的均方位移随时间成正比地增加。实际上,如果在 n 步后分子在某个方向上移动了距离 l_n,则在 $n+1$ 步之后它移动的距离为 $l_{n+1}=l_n\pm\Delta l$。且有 $(l_{n+1})^2=(l_n\pm\Delta l)^2=(l_n)^2\pm2l_n\Delta l+\Delta l^2$。即,平均上,$(l_{n+1})^2=(l_n)^2+\Delta l^2$,因为 $\pm2l_n\Delta l$ 这一项的平均值等于零。因为在时间 t 内分子行进了 t/t_{kinet} 步,则在这段时间内的均方位移为

$$l_t^2=(t/t_{\text{kinet}})\Delta l^2$$

由于 $\Delta l\sim vt_{\text{kinet}}$,则

$$l_t^2\sim(t/t_{\text{kinet}})(vt_{\text{kinet}})^2=t(v^2t_{\text{kinet}})$$

而因为 $t_{\text{kinet}}\approx m/(3\pi D\eta)$,且 $mv^2/2\approx k_BT$,则

$$l_t^2\sim t[k_BT/(1.5\pi D\eta)]$$

我们所得的结果是近似的:因为我们多次使用了符号"≈"("约等于")和"~"("数量级上等于")。但是,正如在这类情形下经常发生的一样,我们得到了相当准确的结果(它需要多得多的精确计算)。精确的结果是这样:

$$l_t^2=t[2k_BT/(\pi D\eta)]\equiv t\cdot[6D_d] \tag{8.20}$$

这里 $D_d=k_BT/(3\pi D\eta)$,为在温度 T 下直径为 D 的球体在黏度系数为 η 的媒质中计算得的爱因斯坦扩散系数值。

特征扩散时间是分子移动与自身尺度 D 相当的距离时所需要的时间。若已知式(8.20),很容易估算这个时间:

$$t_{\text{diff}}=(\pi D^3\eta)/(2k_BT) \tag{8.21}$$

在室温下,由于水的黏度为 $\eta\approx0.01\text{g}/(\text{cm}\cdot\text{s})$,则 $k_BT\approx600\text{cal/mol}$,于是

$$t_{\text{diff}}\approx0.4\times10^{-9}\text{s}\cdot(D/\text{nm})^3 \tag{8.22}$$

式中,(D/nm) 为以纳米为单位的分子直径。

可以证明,分子的翻转(这对应于分子极性在距离 $\approx D$ 上的位移)也需要大致相同的时间。

备注:上述的扩散时间估算对应于水环境。根据现有的数据,蛋白质分子的细胞环境的黏度大约要高一个数量级。因此在那种环境中所有的特征扩散时间大约比纯水环境中要高一个数量级(式 8.21)。

因此,我们可以作出结论,分子在水中的特征扩散时间为纳秒量级:这是分子在经受与环境分子的大量碰撞之后,移动或翻转与自身尺度相当的距离所花的时间。

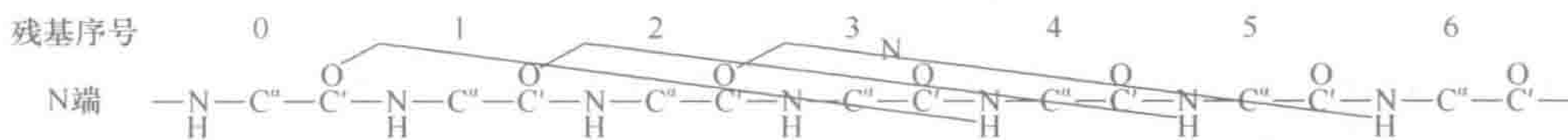
发生得慢得多的过程,让我们可以认为在它的路径中存在有自由能垒。在分子间的反应(或远程链块之间的反应)中,这种能垒可能是参与反应的粒子在靠近时的熵损失所形成的。

第 9 讲

α 螺旋初始化和延长的自由能。朗道理论与螺旋–线团转变的非相变性。螺旋–线团转变中的协同区段的尺寸。α 螺旋在水中的稳定性。β 结构在水中的稳定性。α 螺旋和 β 结构的形成速度。“线团”是什么？

在学习了基础物理学之后，现在我们来讨论二级结构的稳定性及其形成动力学。本讲中我将只讲述**同聚**①多肽——由同一种氨基酸构成的多肽。

■ 让我们从 α 螺旋开始。其中的第一个氢键，$(CO)_0$—$(HN)_4$，固定了 3 个残基（残基 1、残基 2、残基 3）的构象；下一个氢键，$(CO)_1$—$(HN)_5$，只是进一步固定了 1 个残基：残基 4；氢键 $(CO)_2$—$(HN)_6$ 进一步固定了残基 5，依此类推。



这意味着，如果在螺旋中固定了 n 个残基，则它们是由 $n-2$ 个氢键固定的。现在来看在水中由线团（“线团”是没有固定结构、且沿链远离的残基之间没有相互作用的多肽）形成这种螺旋时的自由能。这一自由能可用下式描述：

$$\Delta F_{\alpha}=F_{\alpha}-F_{线团}=(n-2)f_{H}-nTS_{\alpha}=-2f_{H}+n(f_{H}-TS_{\alpha}) \tag{9.1}$$

式中，f_H 为形成氢键的自由能加上随之而在 α 螺旋中出现的相互作用的自由能（你们应该记住，f_H 不只是在真空中的能量：它包含氢键在水中重建时的能量和熵），而 S_{α} 是一个残基被固定到螺旋中时的熵损失。

可见，在 ΔF_{α} 中包含两项，第一项 $(-2f_H)$ 与螺旋长度无关，物理量

$$f_{init}=-2f_{H} \tag{9.2}$$

在传统上被称为螺旋初始化自由能（实际上，考虑到螺旋的起始点和终止点，f_{init} 是螺旋与线团的两个边界的总自由能）。另一项，$n(f_H-TS_{\alpha})$，直接正比于螺旋长度；物理量

$$f_{el}=(f_{H}-TS_{\alpha}) \tag{9.3}$$

被称为螺旋延长一个残基时的自由能。合起来就得到

$$\Delta F_{\alpha}=f_{init}+n\cdot f_{el} \tag{9.4}$$

在这种情形下，由 n 个残基组成的链的纯螺旋状态与其纯线团（彻底地消除了螺旋

①参见第 1 讲“高分子聚合物”的脚注。

成分)状态的概率之比等于

$$\exp(-\Delta F_\alpha/kT)=\exp(-f_{init}/kT)[\exp(-f_{el}/kT)]^n=\sigma s^n \tag{9.5}$$

这里我用了通用的表示方法:

螺旋延长因子:$s=\exp(-f_{el}/kT)$;

螺旋初始化因子:$\sigma=\exp(-f_{init}/kT)$。

显然,$\sigma<<1$,因为 $\sigma=\exp(-f_{init}/kT)=\exp(+2f_H/kT)$;而氢键的自由能为很大的负值,约为几个 kT。

$\exp(-\Delta F_\alpha/kT)=\sigma s^n$ 即为由 n 个残基组成的链在两个状态("α 螺旋"和"线团")之间的平衡常数。

■ 在讲述实验上如何测定 σ 值和 s 值之前,先来解释一个基本的问题:在环境条件(温度、溶剂等)改变时螺旋是怎样形成的:是"全或无"转变,还是渐变?

从表面看来,像螺旋这样的一种与线团不同的结构,应该是通过相变(即"全或无")的方式从线团中"冻结"而成的,就像冰从水中出现一样……

但是对这个问题,朗道理论指出,在两个相都为**一维**的系统中,一级相变是不可能的。我来试着阐述这个理论。

首先,"一维性"的意思是什么?这意味着,两相的"接头"的边界大小以及自由能,与两相各自的大小无关。在这个意义下,多肽中的螺旋和线团都是一维的。图 9-1 表明,螺旋与线团的边界(接头)既与螺旋长度无关,也与线团长度无关。而三维相(如水中的冰)的边界,与表面一样,都与相的大小有关。相应地,螺旋与线团的边界的自由能与它们的长度无关,而三维相的表面自由能随所包含的粒子数 n 的增加而以 $n^{2/3}$ 的速率增加。

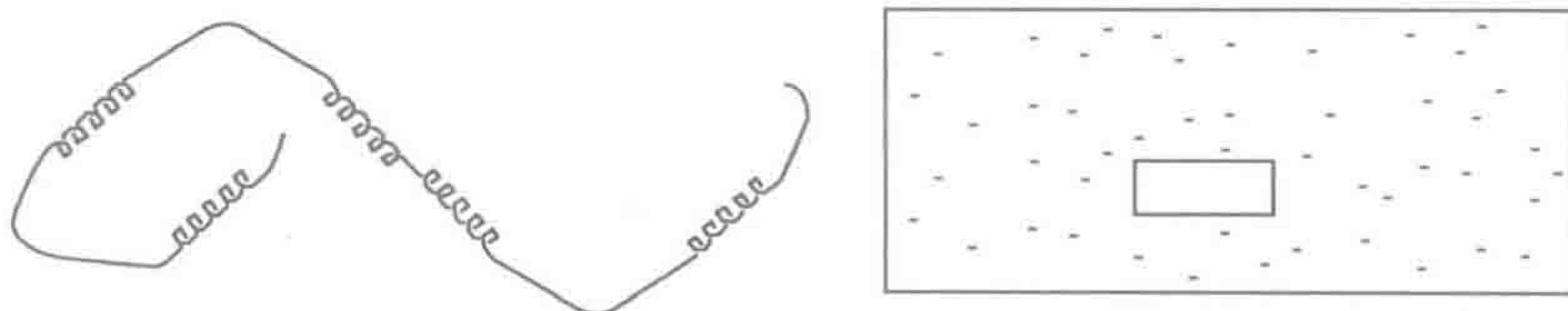


图 9-1 一维系统(含螺旋的线团)与三维系统(水中的冰)的对比。螺旋与线团的接头大小与它们的长度无关;三维冰的表面与其大小有关

■ 现在,来看"通过一级相变而发生"是什么意思。这意味着,在转变温度下,一种相或另一种相都可能是稳定的,但混合相(如冰和水)会导致自由能的升高,因而是不稳定的。这里我们不应该被"浮于水中的冰"的景象所误导:在任何温度下该状态都是不稳定的(由于在水与冰的交界面上增加的自由能),随着时间的继续,冰要么融化,要么占据全部水,当然,如果没有地下热量、水流和其他非平衡物质的流动……

在三维系统中,相的共存是有利的吗?**是不利的**。为什么?请再回到图 9-1。

考虑无限的水和无限的冰具有相同的自由能(这就是“转变点”的条件)时的温度。如果在水中浮有由 n 个分子组成的冰块,边界的自由能正比于 $\xi n^{2/3}$,其中 $n^{2/3}$ 为边界分子的特征数,而每个分子的边界自由能 $\xi>0$(备注:如果 $\xi<0$,“混合”在这时总是热力学有利的,将会在分子尺度发生,根本不可能出现两个相)。这意味着,自由能按 $\xi n^{2/3}$ 随冰块表面而升高。确实,因为冰块可以处于容器中的不同位置,所以它还具有位置熵。但是,如果在容器中共有 N 个分子(即冰可以开始出现的 N 个点),位置熵也不可能超过 $k\ln(N)$ 数量级。总之,冰块的总自由能是 $[\xi n^{2/3}-kT\ln(N)]$ 的数量级。然而,在 N 值很大时,对数增长非常缓慢。如果冰块占据容器中相当大的一部分(例如,$n\sim N/10$),且 N 非常大(例如,10 000 000 000),则 $\ln(N)$(这时为 23)与 $(N/10)^{2/3}$(这时为 1 000 000)相比是非常小的,即在冰块的自由能中 $\xi n^{2/3}$ 项是起主导作用的,而它是阻止冰块形成的……所以在三维系统中宏观相分裂(由少量分子形成的“冰块”不考虑在内:它仅仅是微观的,局部涨落),使得在其中可以发生一级相变。

那么,在一维系统中,相的共存是有利的吗? **答案是,是有利的**。重新来看“转变中点”的温度。在此温度下螺旋和线团具有相同的自由能,即 $f_{el}=0$。螺旋与线团的两个边界的自由能 f_{init} 既与螺旋的大小无关,也与线团的大小无关。在长度为 N 的链中,长度为 n 的螺旋的位置熵等于 $k\ln(N-n)$。漂浮在该链中的螺旋的总自由能为 $f_{init}-kT\ln(N-n)$。在 N 很大时,$\ln(N-n)$ 项在该表达式中**总是**起主导作用的,甚至在 $n\sim0.9N$ 时也是如此;而这个对数项会降低自由能,且促进在线团中产生螺旋(同样的,在螺旋中产生线团)……所以在一维系统中相**不会**分裂,它们趋向于混合。一旦如此,则在链足够长的情况下,一级相变(即“全或无”型相变)**不可能**发生。朗道理论得以证明。

备注:严格地说,与 α 螺旋的熔解不同的是,DNA 双螺旋的熔化不服从朗道理论。因为 DNA **不是**一维系统。在双螺旋段熔化时形成了终端闭合的空间环,而环的这种闭合为双螺旋与环的边界自由能提供了额外的贡献,它随着环长的增加而指数性增加。

现在可以提出一个问题:线团相与螺旋相在什么样的链特征长度下开始混合? 或者说,在螺旋-线团转变中,螺旋片段的特征长度 n_0 是多大?

来看由 N 个残基构成的链在“转变中点”温度下的情况,此时螺旋和线团具有相同的自由能,即 $f_{el}=0$。而且螺旋延长自由能为零(线团也一样),螺旋初始化自由能为 f_{init},在 N 个残基构成的链中的螺旋可能的位置数约为 $N^2/2$(在相同的条件下它可以在任意位置开始和终止,长度不小于三个残基);而螺旋在链中的分布位置,以及它的长度(因为 $f_{el}=0$)都不影响它的自由能。为了得到定性的估算,忽略与主量(公式中的字母)相比要小的量(公式中的数字)。于是有:螺旋的位置数约为 N^2,即它们的熵约为 $k\cdot2\ln(N)$,而在长度为 N 的链中插入一个新的相小段(端点波动的螺旋)的总自由能大约为 $f_{init}-2kT\cdot\ln(N)$。如果这个自由能大于零,则

新相无法插入；如果它小于零，则新相可以插入，而且可以多次进行。这意味着，线团相与螺旋相的混合在块长度 $N \sim n_0$ 时开始，而 n_0 值由公式 $f_{init} - 2kT \cdot \ln(n_0) = 0$ 得到。于是有：螺旋块和线团块在转变中点的特征长度为

$$n_0 = \exp(+f_{init}/2kT) = \sigma^{-1/2} \tag{9.6}$$

在实验中，转变中点（温度）对应着极长多肽链的螺旋度为 50% 时的点（如前所述，多肽的螺旋度通常由圆二色性谱测定；在螺旋度为 50% 时，多肽的圆二色性谱看上去像全线团的多肽和螺旋度为 100% 的多肽的圆二色性谱之和的 1/2）。在该点有 $f_{el} = 0$，即 $s = \exp(-f_{el}/kT) = 1$。

在这一条件（$\sigma = 1$）下测量螺旋性与多肽长度的相关性，可以找出 n_0，在该长度下其螺旋性的平均值只有极长链的螺旋性平均值（即 50%）的 1/4。这个数值估算的证明超出了本课程的范围。你们可以试试自己证明。

最后，如果已知 n_0，则可以计算出 f_{init} 和 σ。对于大多数氨基酸，$n_0 \approx 30$，$f_{init} \approx$ 4kcal/mol，且 $\sigma \approx 0.001$。

现在我们可以求出氢键形成（以及形成氢键时在 α 螺旋中所伴随的所有相互作用）的自由能：根据式（9.2），$f_H = -f_{init}/2 \approx -2$kcal/mol。也可以求出固定一个残基到 α 螺旋中时所减小的构象熵：根据式（9.3），在 $f_{el} = 0$ 时，$TS_\alpha = f_H \approx 2$kcal/mol。

螺旋稳定性的两个参数 f_{el} 和 f_{init} 都与温度相关，但对螺旋稳定性影响比较大的是 f_{el} 对 0 的偏离。原因在于，在由 $\sim n_0$ 个残基构成的螺旋中，这一偏离要乘以很大的值 n_0，而且已经以这种形式出现在螺旋的自由能中。当 $f_{el} n_0/kT$ 值约为+1（更准确的值是：$f_{el} n_0/kT = +2$）时，螺旋性实际上消失了，而当 $f_{el} n_0/kT = -2$ 时，线团实际上消失了。这意味着，在极长（$N \gg n_0$）的多肽链中，螺旋-线团转变发生在 $2/n_0 < f_{el}/kT < 2/n_0$ 的范围内，即（$n_0 \approx 30$ 时）$-0.07 < f_{el}/kT < 0.07$（图 9-2）。这是一个急剧的协同转变的例子，但它**不是**相变（因为它的宽度并不随着链长度的增加而趋向于零）。

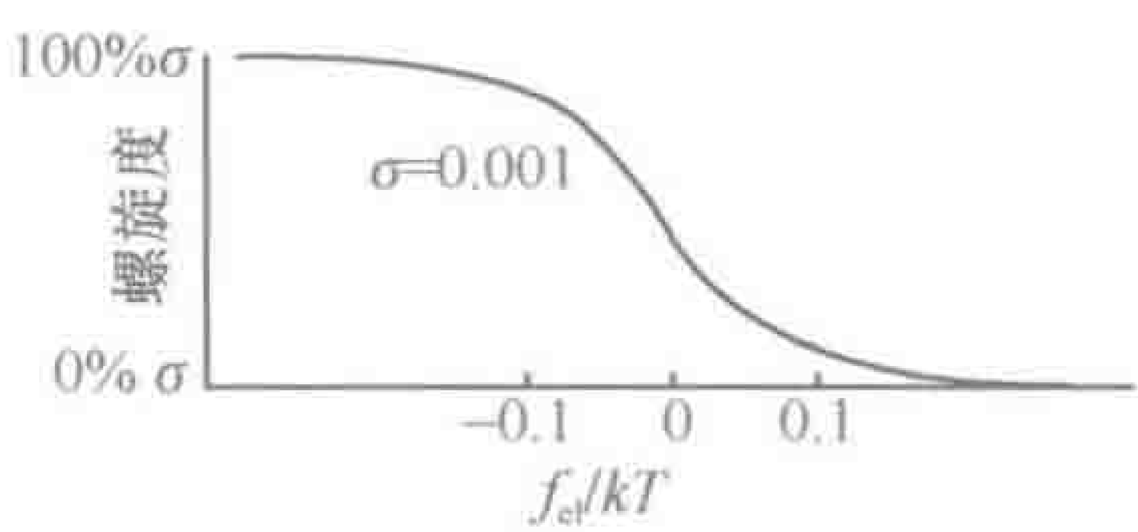


图 9-2 螺旋-线团转变总是（甚至对于无限长的链）具有有限的宽度。这是一个**非相变**、而是协同转变的例子：在螺旋初始化因子值很小（$\sigma \ll 1$）时，它在 f_{el} 值变化很小（远小于 kT）时完成。这表明，“转变单元”涉及链中的许多残基，但远不是整个键

■ α 螺旋的稳定性通常随温度的升高和极性变性剂的添加而降低，随弱极性溶剂的添加而升高。

为了测定单个氨基酸残基对螺旋稳定性的影响，目前主要是使用短链多肽（长度 $\sim n_0$ 或更短）。在它们中只能形成一个螺旋，就非常容易估算每个残基替换对螺旋性的影响。目前已知，氨基酸残基对螺旋稳定性的贡献在如下范围中：丙氨酸（Ala）是最有利于“螺旋形成”的残基：$s \approx 2$，即 $f_{el} \approx -0.4$kcal/mol；甘氨酸（Gly）是利于“螺旋破坏”的残基：$f_{el} \approx +1$kcal/mol，即 $s \approx 0.2$；实际上，脯氨酸（Pro）是**亚**氨基酸（它没有 NH 基团可参与形成 α 螺旋的氢键），它的 σ 值要小得多，为 0.01 ~

0.001，但还没有准确测定。

以前，通过使用统计性共聚物（如以随机方式把80%的Glu和20%的Ala混合而成的链）做过类似的估算（其中有本书作者之一：Oleg B.Ptitsyn），正是通过这种方法得到了最早的（意味着也是最重要的）估算。不过随着能按照预设序列合成多肽的新方法的出现，这种基于随机共聚物的工作看上去已经成为了过去式。

通过使用一级结构已经预先设定的短链多肽，甚至可以估算每一个特定的氨基酸在这种多肽上每一个具体位置上的替换对螺旋性的影响，即实际上是估算氨基酸残基对螺旋稳定性的贡献对残基至螺旋的N和C端位置的依赖关系。侧链基团，尤其是带电基团，与这些末端相互作用各不相同，因为如前所述，在螺旋的N端存在有未参与氢键的主链NH基团（且α螺旋的N端局部带电量等于$+e/2$），而在C端则有游离的CO基团（总局部带电量为$-e/2$，半个电子的电量）。

除此之外，还可以使用电位滴定法测量含有酸性或碱性侧链基团的多肽[如poly（Glu）或poly（Lys）]中的螺旋的稳定性（f_{el}值）。这种方法的原理在于通过使螺旋带电而破坏它（因为在螺旋中侧链基团的电荷比在线团中更接近，因而强烈地排斥），因此如果已知链的总电荷与环境pH的依赖关系，就可计算出螺旋的稳定性。遗憾的是，对这一有趣方法的更详细讨论超出本课程的范围。

类似的方法可以应用于测量多肽中的β结构的稳定性。但是它们发展不多，因为β结构是强烈聚集的。目前β结构的稳定性经常直接在蛋白质中测定：通过估算位于β结构表面的单个氨基酸残基的点替换对它的稳定性的影响。这类替换对β结构的影响在-1~+1kcal/mol。各种残基稳定α和β结构的能力，我们将在以后的课里讨论。

■ 现在来看在多肽中二级结构的形成速率。

α螺旋的形成非常快：已经证明，20~30个残基的多肽在0.1μs内就能形成螺旋构象（这样快速的测量需要皮秒或纳秒级激光照射产生的溶液热变）。意味着，螺旋的延伸速率至少约为几纳秒一个残基。

我说“至少”，是因为螺旋的形成速度不仅与它的延伸速率有关，还与链中出现第一批螺旋结构晶核的速率有关。而螺旋的初始化需要克服活化能垒，因此形成第一圈螺旋比后续延伸一圈需要更多的时间。因此有可能螺旋的延伸非常快，而整个观测时间都花费在螺旋的初始化上。让我们更仔细地考虑这个问题。

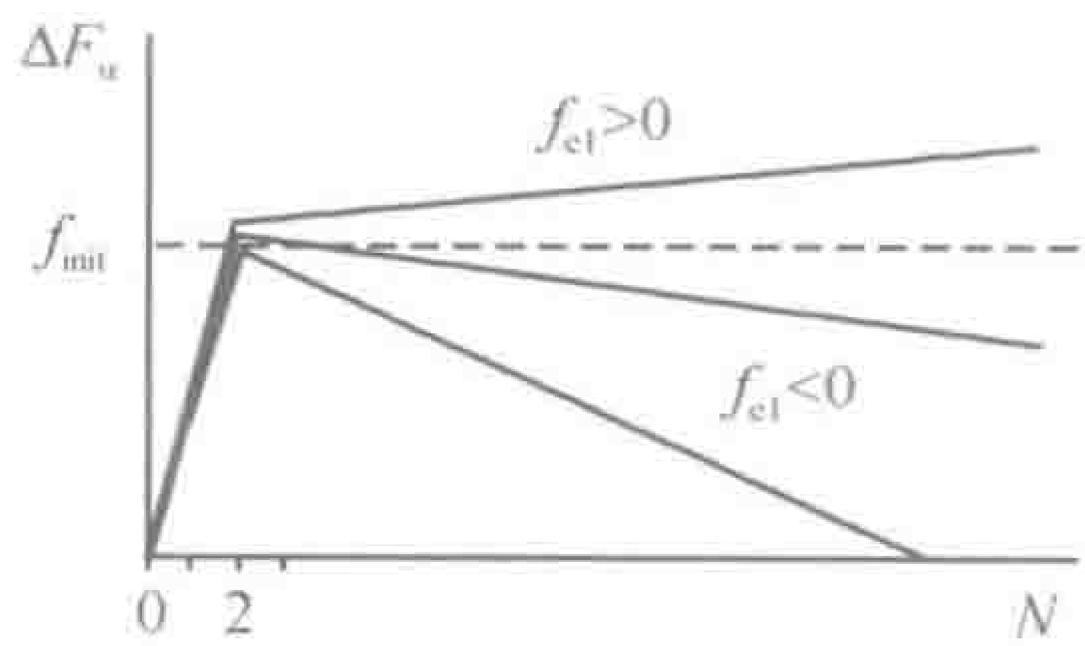


图9-3 在不同的延伸自由能（f_{el}）下，螺旋自由能（ΔF_α）与螺旋中所包含的残基数（N）的典型依赖关系。在$f_{el}<0$时，长螺旋是稳定的，但它的形成需要初始化，为克服高度为f_{init}的活化能垒。在$f_{el}>0$时，任何长度的螺旋都不稳定，因而不能形成

螺旋自由能与其长度的典型依赖关系显示在图9-3中。甚至在$f_{el}<0$时（即相当长的螺旋是稳定的），其中形成第一圈也需

要克服高度为f_{init}的活化能垒。

按照过渡态理论,克服这种能垒并在链上给定位置形成第一圈发生在如下时间内:

$$t_{\text{init_}\alpha}=\tau\exp(+f_{\text{init}}/kT) \tag{9.7}$$

式中,τ 为基元步骤(这里是螺旋延长 1 圈)的时间,而指数是考虑到具有(根据"能垒"的定义!)最高自由能的能垒状态的低占据率,根据式(9.6),有

$$t_{\text{init_}\alpha}=\tau/\sigma \tag{9.8}$$

然而初始化可以发生在未来螺旋的任意点,而它的长度(甚至在极长的肽链中)约为 $n_0=\sigma^{-1/2}$。这意味着,在未来螺旋的**任意位置**的初始化第一圈的特征时间少至 n_0 分之一,而 $t_{\text{init}}/n_0=\tau/(n_0\sigma)=\tau/\sigma^{1/2}$。

然后,已经初始化的螺旋延伸到它所有 $\sim n_0$ 个残基,需要大约相同(与初始化相比)的时间,$\sim\tau n_0=\tau/\sigma^{1/2}$。这意味着,在多肽上的线团-螺旋转变的总时间应该约为 $2\tau/\sigma^{1/2}$,而半转变的特征时间为

$$t_\alpha\sim\tau/\sigma^{1/2}=\tau\exp(+f_{\text{init}}/2kT) \tag{9.9}$$

粗略地说,一半的时间花在链上任意位置的螺旋初始化上,而另一半花在延伸上,根据前面的估算,速度约为每纳秒一个残基。

■ α 螺旋的形成动力学相当简单:它们总是快速形成。而 β 结构的形成动力学复杂得多,也更有趣。

多肽中的 β 结构通常形成得很慢,几小时,甚至几天(虽然也有在几毫秒内的)。什么原因呢?而且,含 β 结构的蛋白质折叠并不比 α 螺旋形的蛋白质更慢(在任何情况下,并不慢很多)。它们是如何做到的?是什么引起了多肽中的 β 结构形成速率的异常:初始化慢还是延伸慢?

■ β 结构形成动力学"异常"(与螺旋-线团转变相比)的原因在于,β 结构是二维(而不是像螺旋或线团那样是一维的)对象(图 9-4)。因此,β 结构的形成可以像一级相变那样进行。让我们来详细分析这一过程。

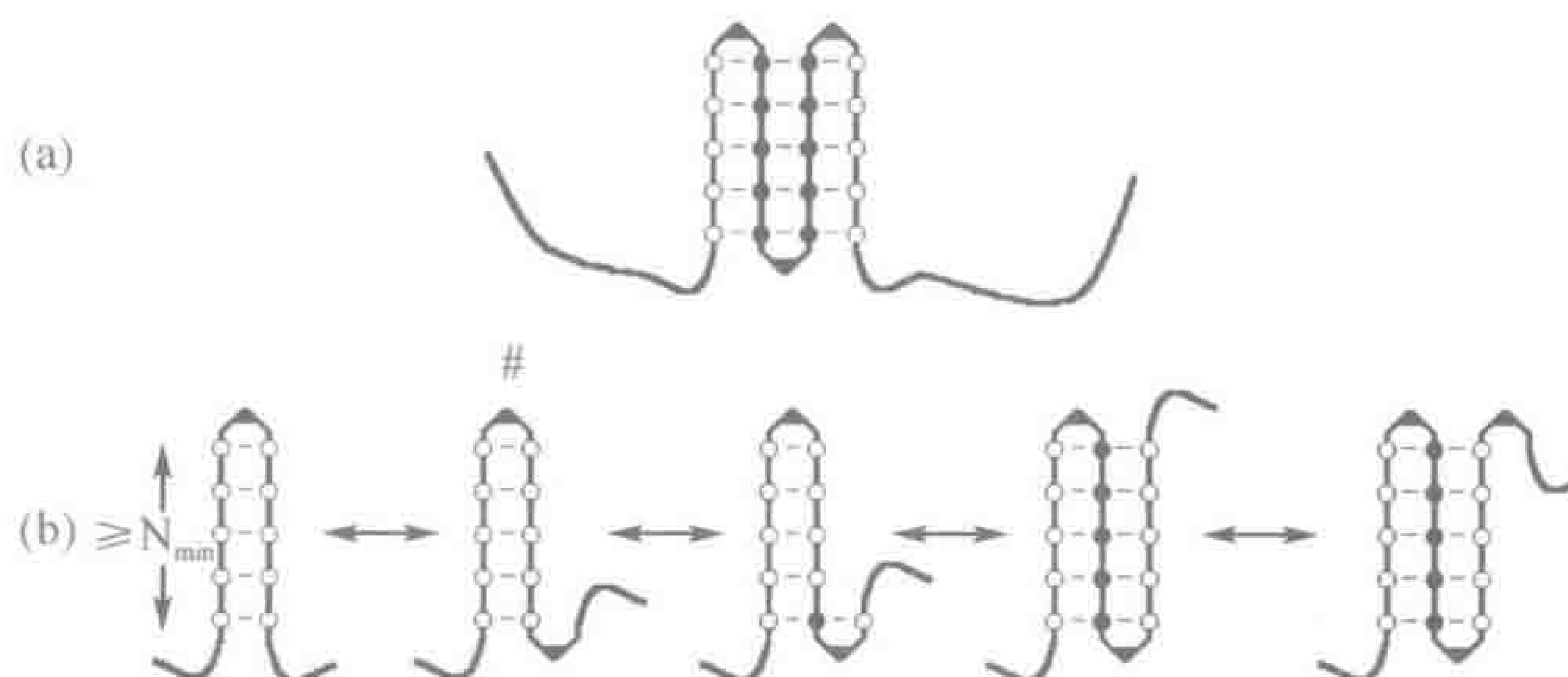


图 9-4 (a)β 片示意图。内部 β 片段的氨基酸残基以实心圆表示,边缘的显示为空心圆;连接 β 片段的转折(或环线)以拐角表示;(b)正文中所描述的 β 片生长的次序(单独的 β 发夹不稳定的情形)。最不稳定的结构上标记了#号

处于β片边缘的残基与其他残基的连接数比处于β片内部的残基要少。换句话说,β片的边缘(正如任意其他相一样,如水滴、冰块或α螺旋)有较高的自由能。然而,与α螺旋不同(而与水滴相似)的是,β片不是一维的,即它的边界(也意味着,它的自由能)随着参与其中的残基数的增多而增大。因此,线团-β结构转变是一级相变,就像水滴或冰块的形成一样。可以证明,因此在β片的形成路径上(特别是在形成轻微稳定的β结构时)出现会高自由能的能垒,可使折叠初始化变慢到百万倍或十亿倍。

β片的边缘是由①边缘β串和②连接β串的转折或环线构成的[图9-4(a)]。设线团的自由能为0(即参考点),f_β 为在β片中央的残基的自由能,$f_\beta+\Delta f_\beta$ 为边缘处残基的自由能(即 Δf_β 为边缘效应),U 为转折的自由能。β片一旦完全形成,则是稳定的(即 $f_\beta<0$),而边缘效应阻止它破裂成小块(即 $\Delta f_\beta>0$ 且 $U>0$)。

在β结构形成动力学中必须分清两种情形。

(1) $f_\beta+\Delta f_\beta<0$,即长β发夹本身比线团更稳定。则仅在该发夹顶部形成转折时需要克服活化能垒(大致与α螺旋的形成相似),而β结构的继续生长进行得很快,大致与α螺旋的延伸相同(图9-3中 $f_{el}<0$ 的直线)。

(2) $f_\beta+\Delta f_\beta>0$,即β发夹本身是不稳定的,只有初始的β发夹上连接另一个β串才能够使β片稳定。则活化能垒对应于形成这种β片或β发夹的"晶核",β片的继续生长伴随着自由能的总体下降。

■ 新相的晶核的形成和随后的"交互生长"是一级相变的典型特征,β结构的形成就属于此类。然而,克服由形成晶核所引发的活化能垒,正如我们很快要证明的,可能会非常非常慢。

来看下述最简单的形成稳定β片的顺序[图9-4(b)]:由一个转折和两个各含 N 个氨基酸残基的β串形成初始的β发夹;在它的末端形成下一个转折;黏接另一个含 N 个残基的β串;形成下一个转折;黏接β串;如此继续。

从线团中生成由一个转折和两个各含 N 个残基的β串构成的β发夹,使链的自由能升高 $U+2N(f_\beta+\Delta f_\beta)$;形成下一个β转折继续使自由能增加 U。在这个发夹的边缘黏接一个由 N 个残基构成的β串使自由能降低 Nf_β(因为边缘的残基数没有变化,而内部的残基数增加了 N);形成下一个β转折再次使自由能增加 U;黏接另一个由 N 个残基的β串再次使自由能降低 Nf_β,依此类推。

"连接新β串并形成新β转折"的循环使总自由能改变 $Nf_\beta+U$。由于这种循环必须使自由能降低(只有这样才能发生链的折叠),所以黏接的β串(以及与之黏接的β串)应该至少包含 $N_{min}=U/(-f_\beta)$ 个残基。

"过渡态",即β结构形成过程中最不稳定的状态,是(按照上述顺序)带有下一个转折的β发夹。由于发夹的稳定性随着长度的增加而降低,而初始发夹中的β串应该至少包含 N_{min} 个残基,所以初始β发夹(带有下一个转折)的最小自由能为

$$F^{\#}=U+2N_{\min}(f_\beta+\Delta f_\beta)+U=2(U\Delta f_\beta)/(-f_\beta) \qquad (9.10)$$

这就是**按照上述顺序**形成β片的过程中的过渡态自由能。现在只需证明，无论采取什么顺序，都不可能存在更稳定的过渡态。

为了证明这一点，试估算β片的最小自由能如何随它的增大而变化。自由能变化的一般过程显示在图9-5中。当β片较小时，边缘效应起主导作用，自由能升高。在β片很大时，内部残基起主导作用，其自由能降低。

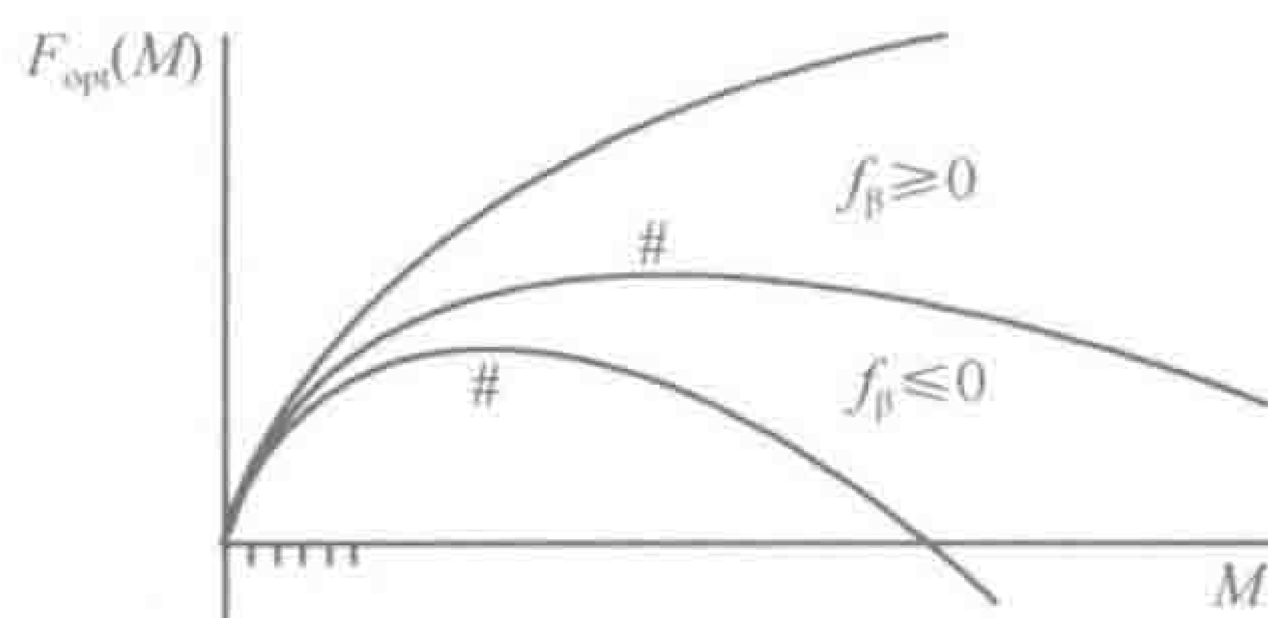


图9-5　β片的最小自由能 F_{opt} 随链中的残基数 M 增大而变化的一般过程。各条曲线对应于不同的β片内部残基自由能 f_β。#处为 F_{opt} 在β片增大过程中的最大值。在β片增大时(与α螺旋增大不同，图9-3)其自由能最大值并不是出现在该过程的最初开始处

现在来进行计算。首先来估算含 M 个残基的β片的最小自由能。由 m 个β串(长度相同，以便使边缘自由能最小)和 $m-1$ 个转折构成的β片的自由能为

$$F(M,m)=Mf_\beta+2\frac{M}{m}\Delta f_\beta+(m-1)U \qquad (9.11)$$

(转折中的残基数较少，此处忽略了)。

改变β串的数目 m(对于给定的 M)，可以由如下条件得到该自由能的最小值：

$$\frac{\mathrm{d}F}{\mathrm{d}m}=-2\frac{M}{m^2}\Delta f_\beta+U=0 \qquad (9.12)$$

由此得到在该β片中的β串的最佳数目 $m_{opt}=\sqrt{M}\times\sqrt{\frac{2\Delta f_\beta}{U}}$，以及其最佳自由能 $F_{opt}(M)=F(M,m_{opt})=Mf_\beta-U+2\sqrt{M}\times\sqrt{\frac{2\Delta f_\beta}{U}}$。对 M 改变 $F_{opt}(M)$，由如下条件得到其最大值(图9-5)：

$$\frac{\mathrm{d}F_{opt}}{\mathrm{d}M}=f_\beta+\sqrt{M}\times\sqrt{2\Delta f_\beta U}=0 \qquad (9.13)$$

然后可以确定对应于这个最大值的β片的大小为 $M^*=\frac{2\Delta f_\beta U}{(-f_\beta)^2}$，以及其自由能：

$$F^*=F_{opt}(M^*)=\frac{2\Delta f_\beta U}{-f_\beta}-U \qquad (9.14)$$

F^* 和 $F^{\#}$[式(9.10)]的主项相同,都是 $2(\Delta f_\beta U)/(-f_\beta)$。正是由于该项的作用,在使 β 结构稳定的自由能($-f_\beta$)很小时,即 β 结构处于稳定性的边缘时,过渡态的自由能 $F^{\#}$ **总是**较大。实际上,当 β 结构接近于与线团达成热力学平衡时,即当($-f_\beta$)→0 时,$F^{\#}$趋向于无限大。

■ 上面我们讨论了 β 片的折叠初始化自由能。根据过渡态理论,过程(这里为折叠初始化)的时间依赖于过渡态(最不稳定的中间态)的自由能 $F^{\#}$,为

$$t_{\text{init_}\beta} \sim \tau_\beta \exp(+F^{\#}/kT) \tag{9.15}$$

式中,τ_β 为基元步骤(这里是 β 片延伸一个残基;且没有理由认为 β 片的延伸速率与螺旋延伸速率有本质差别)的时间。

$t_{\text{init_}\beta}$为在链上**某个**位置初始化形成 β 片的特征时间。在由 M 个残基构成的链上的**任意**位置初始化形成 β 片的特征时间等于 $t_{\text{init_}\beta}/M$。该 β 片扩展到整个链所需时间约为 $M\tau_\beta$。在短链上(有 $t_{\text{init_}\beta}/M>M\tau_\beta$),初始化过程限制了折叠速率。在这种情况下形成 β 结构的特征时间约为 $\tau_\beta \exp(+F^{\#}/kT)/M$。

对于在长链中(和/或分子间)的 β 结构,特征形成时间约为 $\tau_\beta \exp(+F^{\#}/2kT)$。[这一问题与先前已解答的估算在长链上形成 α 螺旋的时间的问题相似。基本的解答思路是,单个 β 片包含的残基数 M_{eff},其延伸时间与该 β 片的初始化时间大致相同,即 $\tau_\beta \exp(+F^{\#}/kT)/M_{\text{eff}} \sim M_{\text{eff}}\tau_\beta$。]

因此,考虑到 $F^{\#}=2(U\Delta f_\beta)/(-f_\beta)$[式(9.10)],在两种情况下,当$-f_\beta$ 很小时,β 结构的特征形成时间与每残基的稳定度($-f_\beta$)的相关曲线具有如下形式:

$$t_\beta \sim \exp[A/(-f_\beta)] \tag{9.16}$$

无论常数 A 的数值如何,可以看出(图 9-6),在$-f_\beta$ 很小时,β 结构的形成时间都会很大(极限是无限大)。

实验也证明了,正如图 9-6 所示,β 结构的形成速率与其稳定度强烈相关。

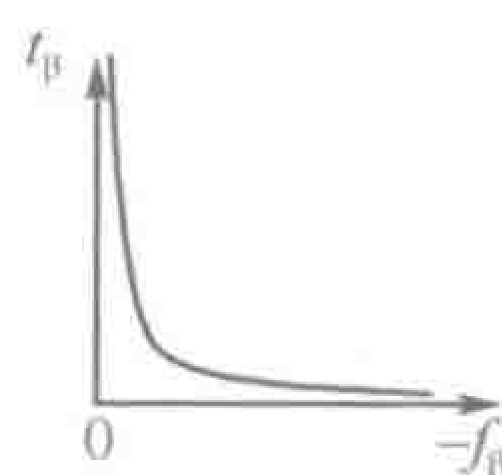


图 9-6 β 片的特征形成时间与 β 结构稳定度的关系曲线的一般形状

总之:稳定度低的 β 结构必定形成得非常慢,且原因不在于延伸缓慢,而在于初始化缓慢:从热运动中积累自由能并克服很高的能垒很困难。

这也解释了实验上观察到的在非聚集性多肽(其中的 β 结构稳定度总是很低)中的 β 结构形成速率极低,而这一速率在 β 结构稳定度提高时迅速上升。

然而,稳定的 β 片和发夹(而在蛋白质中被观察到的正是这些稳定结构)应该形成得相当快,几乎与 α 螺旋一样。

■ 在一级相变中,当新生相处于稳定性边缘时,极其缓慢的初始化是这类相变的基本性质。回忆一下:过冷液体,过冷蒸汽……所有这些都与两相之间的宽阔交界

面上的高自由能有关。既然β结构是以一级相变的方式形成的,则必然具有同样的结果……

相反,α螺旋**不是**以一级相变的方式形成的[回忆一下:与β结构(或冰)不同的是,螺旋的边界**不会**随着螺旋的大小而增大],因此在螺旋折叠时所需要克服的能垒总是具有有限的(且不大的)值,从而可以在几分之一毫秒内完成初始化。

■ 最后的备注:"线团"是什么。这个术语被我多次使用,指的是不具有固定结构的链。其实,"线团"这一概念包括数量极多的不具有任何沿链长程秩序的构象,但是,也不排除微弱的短程(只有几个连续残基的尺度)秩序。

线团最有趣的特性(实验上通过在光散射和X射线散射下的流体动力学方法而观察到)是它密度极低,体积很大,且其大小、半径和体积对链长度的特殊依赖性。

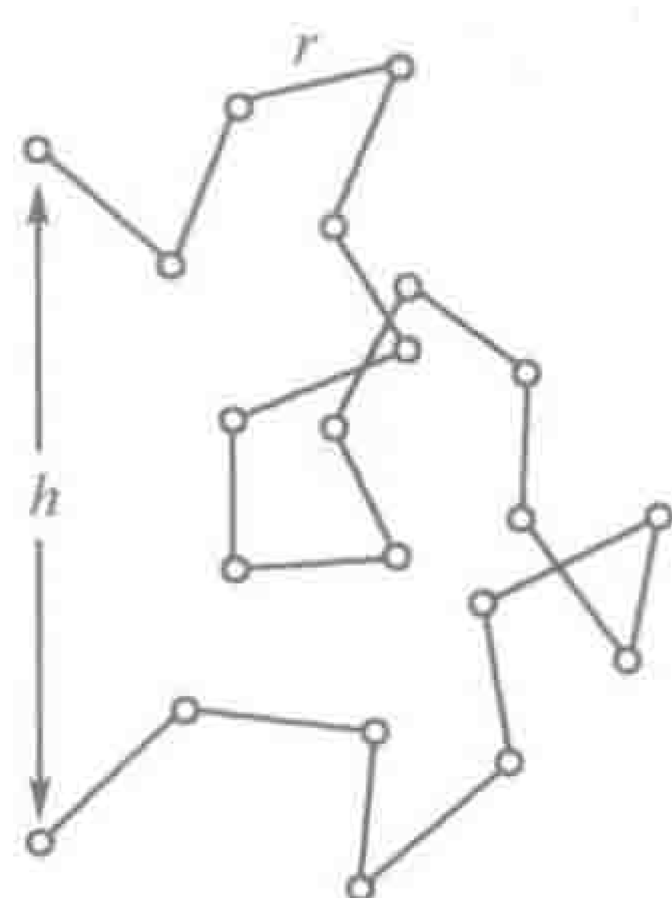


图 9-7 松散联结链:线团模型

为了阐明这一特性,请看一个最简单的被称为"松散联结链"的线团模型(图 9-7)。它是由"小棍"链节(如果链是刚性的,这种"链节"可能对应着几个多肽链单体)构成的;主要特点是,每一个"小棍"可以自由、随意地相对于链上的相邻链节在接头翻转。设链上有 M 个"小棍",每一个的长度为 r。

这种链可以描述为由矢量 $r_1, r_2, \cdots, r_M$(提醒一下:矢量是具有方向的线段)组成的序列。这些矢量具有相同的长度 r,每一个都从前一个接头指向后一个接头。矢量之和为

$$h = \sum_{i=1}^{M} r_i \qquad (9.17)$$

它就是从链的开头指向终点的向量。这一矢量的长度平方为

$$h^2 = \left(\sum_{i=1}^{M} r_i\right)^2 = \sum_{i=1}^{M} r_i^2 + \sum_{i\neq j}^{M}\sum^{M} r_i r_j \qquad (9.18)$$

现在来求 h^2 的平均值 $\langle h^2 \rangle$ 即 h^2 对链构象所有可能的热涨落的平均值(这种平均用符号$\langle\ \rangle$表示)。对此我们应该注意,向量 r_i 的平方平均值$\langle r_i{}^2 \rangle$就等于 r^2,而任意向量的乘积的平均值$\langle r_i r_j \rangle$在 $i \neq j$ 时等于零,因为由于这些向量可以自由旋转,故在各个不同方向上有着相同的概率。

这意味着,松散联结链的两端之间距离的平方平均值等于

$$\langle h^2 \rangle = \left\langle \left(\sum_{i=1}^{M} r_i\right)^2 \right\rangle = \left\langle \sum_{i=1}^{M} r_i^2 + \sum_{i\neq j}^{M}\sum^{M} r_i r_j \right\rangle = \sum_{i=1}^{M} \langle r_i^2 \rangle + \sum_{i\neq j}^{M}\sum^{M} \langle r_i r_j \rangle = Mr^2 \qquad (9.19)$$

即线团的线性长度(半径之类的)随链中的链节数 M 的增加按 $M^{1/2}$ 而增大。由此推得,线团的体积正比于 $M^{3/2}$。而对所有"正常的"(具有固定密度)物体,体积只

按粒子数 M 而增大,即比 $M^{3/2}$ 慢得多。线团的体积对链长的这种异常强烈的依赖性是线团的最突出的性质。特别是,这一性质使得由长链形成的线团的密度极低,而且,结果就是在沿链远程的链节之间实际上不存在接触。

补充。由于线团的体积约为 $M^{3/2}$,而链末端相碰的概率反比于线团占据的体积,故这种相碰的概率正比于 $M^{-3/2}$。这意味着,线团形式的封闭链环的自由能随其长度按 $kT\times(3/2)\ln(M)$(Flory 定律)增长。

■　自由联结模型的不够现实性在于,它允许链上任意链节能够以任意角度转动,而一般来说,实际上不是这样的。但是式(9.19)可以用更普适化的形式改写为

$$\langle h^2 \rangle = Lr \tag{9.20}$$

式中,$L=Mr$,是链的全长("轮廓长度"),正比于链中的链节数,而 r 为链的"松散接头"间的等效距离,即链"忘记"自身方向的特征长度(供参考:如果多肽链中不含太软的甘氨酸和太硬的脯氨酸,则这个特征长度(或者,如常说的"库恩①片段的长度")为 30~40Å,即,这种链的库恩片段约含 10 个氨基酸残基)。

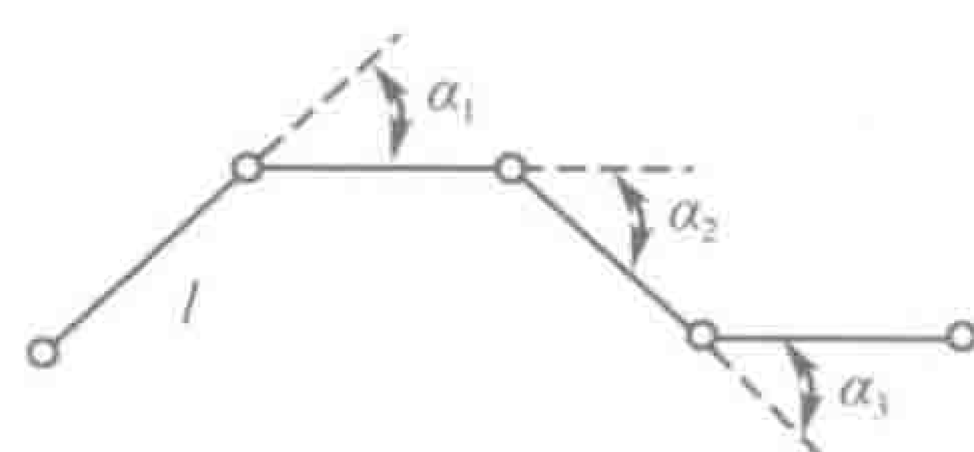


图 9-8　线团的旋转异构模型

描述线团大小的式(9.20)已经具有足够的普适性,可用于描述实际的多聚物。根据该公式,如果 l 为"接头"(链发生转折的点)之间的距离(即共价键的键长),而 $\langle\cos\alpha\rangle$ 为转角(图 9-8)的余弦平均值,则有:

$$r = \frac{1+\langle\cos\alpha\rangle}{1-\langle\cos\alpha\rangle}l \tag{9.21}$$

①理查德·库恩(Richard Kuhn),1900—1967,生于奥地利的德国化学家,研究胡萝卜素、核黄素和维生素成就显著,获 1938 年诺贝尔化学奖。

第10讲

氨基酸残基侧链基团的性质。氨基酸残基参与到二级结构中。丙氨酸(Ala)、甘氨酸(Gly)、脯氨酸(Pro)、缬氨酸(Val)。非极性的、短极性和长极性的侧链基团。带电侧链基团。蛋白质二级结构的疏水性表面。

这一讲讨论氨基酸残基侧链基团的性质。特别地,我想讨论什么样的结构能使各个残基稳定的问题。

■ 20种“标准”(由DNA编码的)氨基酸残基列于表10-1中。表中同时给出了它们的分子质量和在蛋白质中的出现频率。这些氨基酸残基的结构显示于图10-1中。

表10-1 氨基酸残基的基本性质

名称	氨基酸残基			pH7时的分子质量*/Da	25℃下侧链基团的 $\Delta G_{水\to乙醇}$/(kcal/mol)
	缩写符号		在大肠杆菌蛋白质中的百分比/%		
	3字母	1字母			
甘氨酸	Gly	G	8	57	0.0
丙氨酸	Ala	A	13	71	-0.4
脯氨酸	Pro	P	5	97	-1.0
谷氨酸	Glu	E	≈6	128	+0.9
谷氨酰胺	Gln	Q	≈5	128	+0.3
天冬氨酸	Asp	D	≈5	114	+1.1
天冬酰胺	Asn	N	≈5	114	+0.8
丝氨酸	Ser	S	6	87	+0.1
组氨酸	His	H	1	137	-0.2
赖氨酸	Lys	K	7	129	+1.5
精氨酸	Arg	R	5	157	+1.5
苏氨酸	Thr	T	5	101	-0.3
缬氨酸	Val	V	6	99	-1.6
异亮氨酸	Ile	I	4	113	-2.4

* 道尔顿(Dalton,Da,D)是原子质量单位,定义为碳12原子质量的1/12。分子质量即组成分子的所有原子的原子质量。大分子的分子质量通常使用kDa。

续表

名称	氨基酸残基			pH7 时的分子质量*/Da	25℃下侧链基团的 $\Delta G_{水\to乙醇}$/(kcal/mol)
	缩写符号		在大肠杆菌蛋白质中的百分比/%		
	3 字母	1 字母			
亮氨酸	Leu	L	8	113	-2.3
蛋氨酸	Met	M	4	131	-1.6
苯丙氨酸	Phe	F	3	147	-2.4
酪氨酸	Tyr	Y	2	163	-1.3
半胱氨酸	Cys	C	2	103	-2.1
色氨酸	Trp	W	1	186	-3.0

注:疏水性的数据($\Delta G_{水\to乙醇}$)来源于 Fauchére J. L. ,Pliska.V. 1983.Eur J Med Chem Chim Ther.18:369.其他所有数据来源于[3]。残基(在蛋白质或水中)所占据的体积(Å^3)约为其分子质量(Da)乘以 1.3。更确切地说,如果在残基中含有大量脂肪族基团(—CH_2—,—CH_3),则比"分子质量×1.3"约高 5%;如果在残基中含有大量极性原子(O,N),则比"分子质量×1.3"约低 5%。

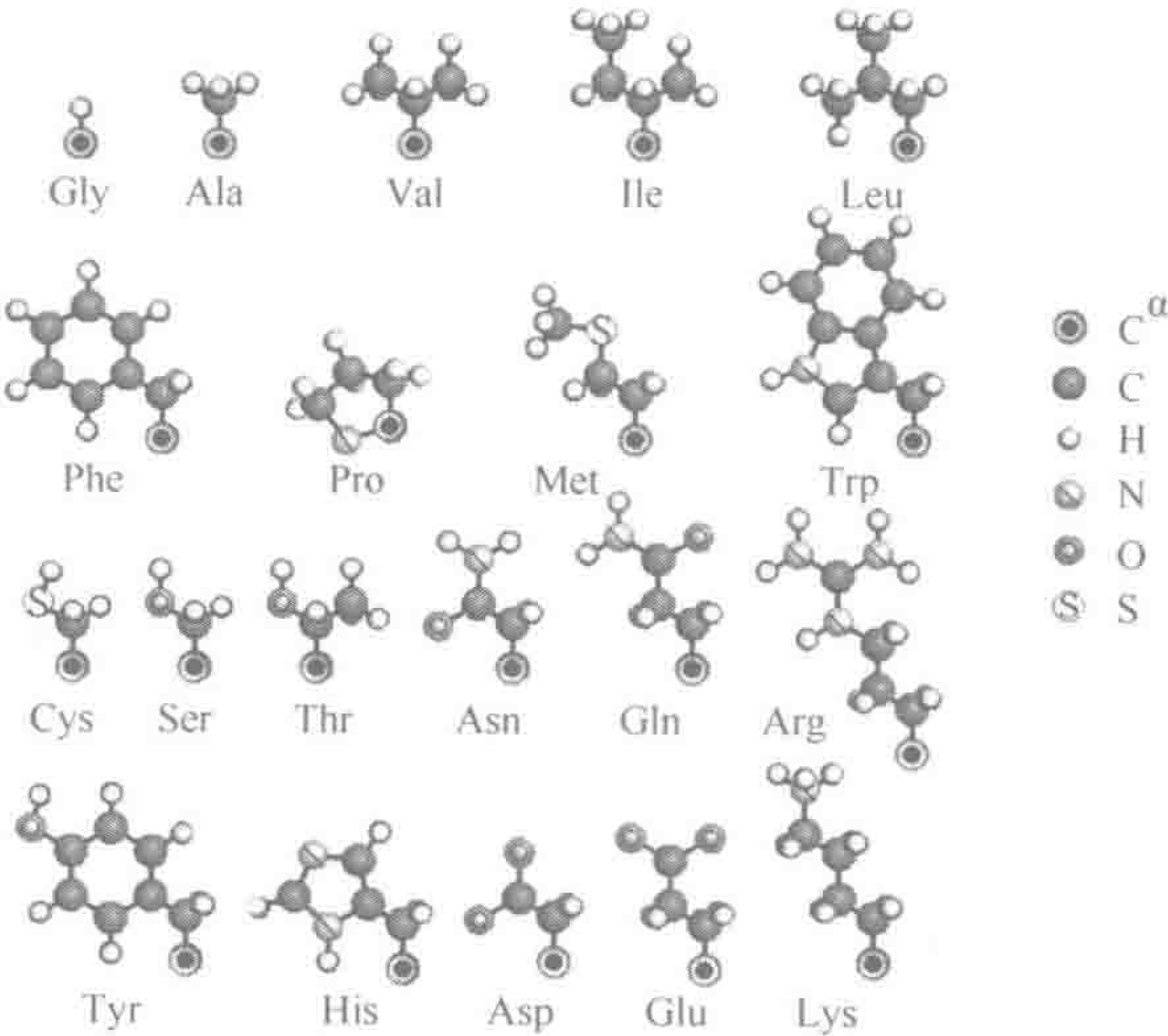


图 10-1　20 种标准氨基酸残基的侧链。图例为组成氨基酸的原子

后来知道,DNA 有时还会编码另一个"非标准的"第 21 种氨基酸:硒代半胱氨酸,即硫原子被硒(Se)所取代的半胱氨酸。硒代半胱氨酸对应于基质 DNA 中的 UGA 密码子,但并非总是如此(通常 UGA 是作为终止密码子用的),只有在特定的核苷酸环境才会这样。而载有硒代半胱氨酸的 tRNA 只认识这种密码子。

■　现在让我们来看氨基酸残基的结构倾向性;它们是在多年的蛋白质结构统计研究之后才弄清楚的。这些研究要回答的问题是:哪些是经常出现的,哪些是不可能出现的。

表 10-2 也许有助于把答案系统化。在表中，除了列出氨基酸残基在蛋白质各种不同位置的出现频率之外，还同时列出了残基的其他性质，例如：主链中是否含有 NH 基团（仅在脯氨酸中无此基团），侧链中是否含有 C^{β} 原子（仅在甘氨酸中无此原子）；侧链中非氢 γ 原子的数目，以及侧链中是否存在极化基团及其类型（偶极子或电荷加正负号表示；对应于“正常的”pH 7.0 的带电状态用粗体标示）。

表 10-2　氨基酸残基的基本结构性质

残基	是否包含		侧链中的 γ 原子数	偶极子/电荷	p*K*	倾向于出现在							
	主链 NH	侧链 C^{β}				α_N 前	螺旋中 α_N	螺旋中 α	螺旋中 α_C	α_C 后	β 中	环线	核中
Gly	+	−						−			−	+	
Ala	+	+						+				−	
Pro	−	+	1				+	−	−	−	−	+	
Glu	+	+	1	COOH⇒$\mathbf{CO_2^-}$	4.3	+	+		−	−	−		−
Asp	+	+	1	COOH⇒$\mathbf{CO_2^-}$	3.9	+	+	−	−	−	−	+	−
Gln	+	+	1	$OCNH_2$									−
Asn	+	+	1	$OCNH_2$		+		−		+	−	+	−
Ser	+	+	1	OH		+						+	−
His	+	+	1	NH；N⇒NH^+	6.5		−		+	+			
Lys	+	+	1	NH_2⇒$\mathbf{NH_3^+}$	10.5	−	−		+	+	−		−
Arg	+	+	1	$\mathbf{HNC(NH_2)_2^+}$	12.5	−	−		+	+	−	+	−
Thr	+	+	2	OH		+					+		
Ile	+	+	2								+	−	+
Val	+	+	2								+	−	+
Leu	+	+	1					+			+	−	+
Met	+	+	1					+			+	−	+
Phe	+	+	1								+	−	+
Tyr	+	+	1	**OH**⇒O^-	10.1			−			+		+
Cys	+	+	1	**SH**⇒S^-	9.2			−			+		+
Trp	+	+	1	NH							+		+

注：在“中性的”pH7 时的带电状态，在表中的“偶极子/电荷”栏中用粗体标示；p*K* 是基团以相等的概率处于带电状态和非带电状态时的 pH（图 10-5 下方）。“结构性质”包含：在 α 螺旋中的出现倾向，尤其是在 N 和 C 端末圈（α_N，α_C），以及直接位于 N 端之前和 C 端之后（α_N 前，α_C 后）；β 结构中的出现倾向（β），不规则结构如“环线”（此处包括 β 转折）的出现倾向；以及在球形疏水核中的出现倾向，但不含在表面的倾向。“倾向”是根据残基在某一给定结构中的浓度相对于该残基在被研究的蛋白质中的平均浓度而量度的。倾向“出现”用“+”表示，“不出现”用“−”表示。用“+”和“−”表示的出现频率之间的差别大约为两倍。更强的倾向用粗体字表示。

试从我们已经学过的知识出发，来理解这个表中的基本规律。在此我们将使用下面的逻辑：因为蛋白质在整体上是稳定的，它应该主要是由稳定的元件组成

的,即这些稳定元件应该在它的结构中更常见,而不稳定的应该比较罕见。

为什么脯氨酸(Pro)不喜欢二级结构？因为它的主链中没有 NH 基团,即它结合成氢键(而二级结构正是由氢键维持的)的可能性减半。然而为什么它喜欢螺旋的 N 端？因为在 N 端,NH 基团是从螺旋中"伸出",即同样不参与氢键,脯氨酸在这里也没损失什么……在另一方面,脯氨酸中的 ϕ 角被它的环固定在大约 $-60°$,即它的构象几乎已经为 α 螺旋"做好了准备"[图 10-2(a)]。

为什么甘氨酸(Gly)不喜欢二级结构,而偏爱不规则片段("线团")？因为在它的拉氏图上角度($\phi\psi$)允许区很宽阔[图 10-2(b)],它很容易采取二级结构之外的各种构象。

相反,丙氨酸(Ala)在拉氏图上的允许构象区窄得多,但是包含了 α 和 β 构象[图 10-2(b)],从而偏爱参与 α 螺旋(和部分 β 结构),而不喜欢不规则构象。

其余的疏水残基(即侧链上不带电荷和偶极子的残基)通常偏爱 β 结构。为什么？因为它们的大 γ 原子在那里更自由[图 10-2(c)]。这对带有两个大 γ 原子的侧链基团尤为重要。

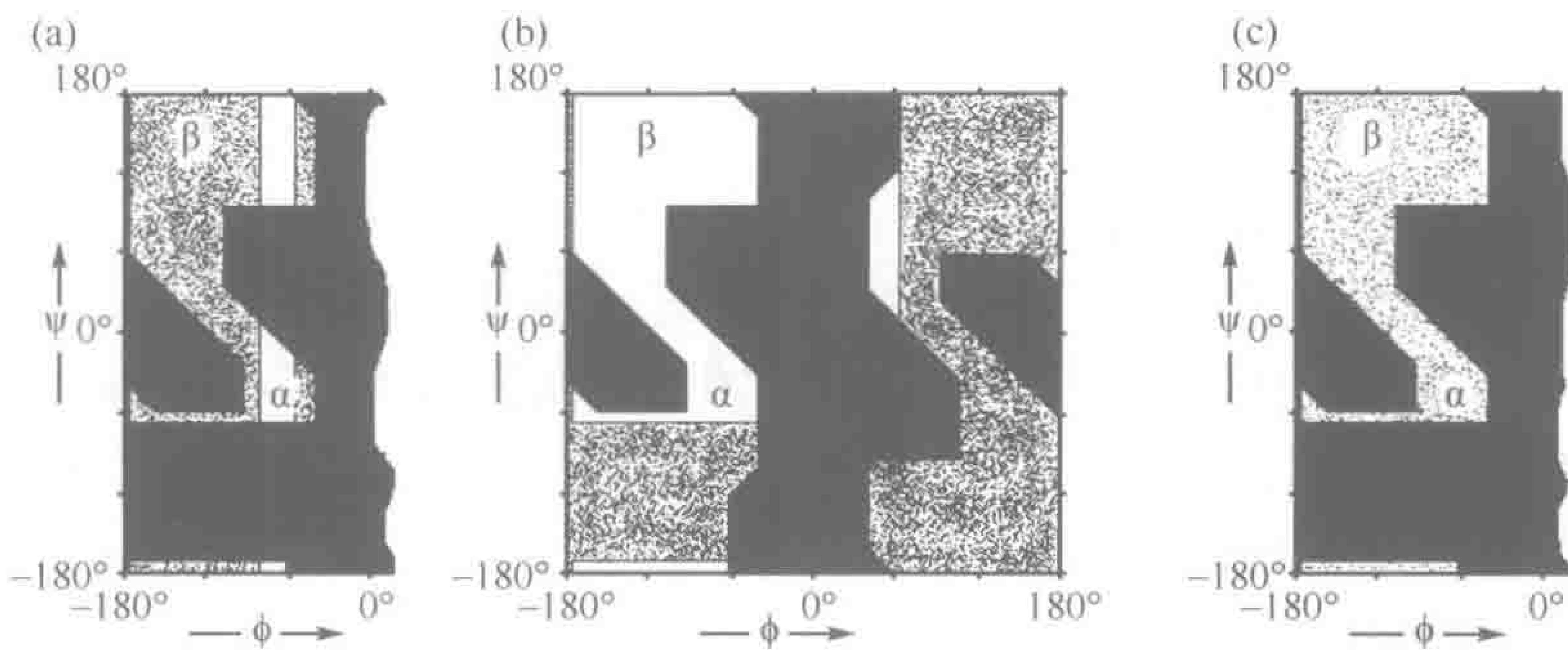


图 10-2　不同氨基酸残基的禁止构象和允许构象,以及以其为背景的 α 和 β 结构的构象。(a)在丙氨酸(Ala)允许构象(▒)背景上的脯氨酸(Pro)的允许构象(□);█为两者都禁止的构象。(b)在甘氨酸(Gly)允许构象(▒)背景上的丙氨酸(Ala)的允许构象(□);█为两者都禁止的构象。(c)更大的残基的禁止构象(█)和允许构象(□,▒)。在□区域是对所有侧链 χ1 角都允许的构象,在▒区域有部分 χ_1 角被禁止

而侧链上含有极性基团的氨基酸,更偏爱不规则片段("线团")的表面,在那里这些极性基团可以与多肽链或水结合成氢键。对于在"中性"pH7 时极性化、带电性最大的残基,这个趋势尤其明显。以极性基团很靠近主链的短极性侧链(图 10-1)也是如此。顺便说到,根据同样的原因,由于在那里可能形成额外的氢键,短极性侧基团喜欢螺旋两端的位置。

侧链中含有偶极子的氨基酸中有两个例外:色氨酸(Trp)和酪氨酸(Tyr),它们在很大的疏水部分的背景之上有小偶极子,半胱氨酸(Cys)也是例外,它的(即其 SH—基团的)氢键非常弱。总之,它们的行为更像疏水残基。

我们还看到，带负电的侧链更偏爱螺旋的 N 端（**更准确点**：N 端的螺旋圈和前面的一两个残基）而且不喜欢螺旋圈的 C 端（及之后的两个残基），而带正电的侧链则更喜欢螺旋的 C 端而不喜欢 N 端。为什么呢？因为在螺旋的 N 端伸出 NH 基团而且在上面出现明显的正电荷，侧链的“负极”与它相吸引，而“正极”与它相排斥（图 10-3）。相反的是，螺旋的 C 端带负电，其效果是相反的：C 端附近喜欢吸引侧链的“正极”，排斥其“负极”。

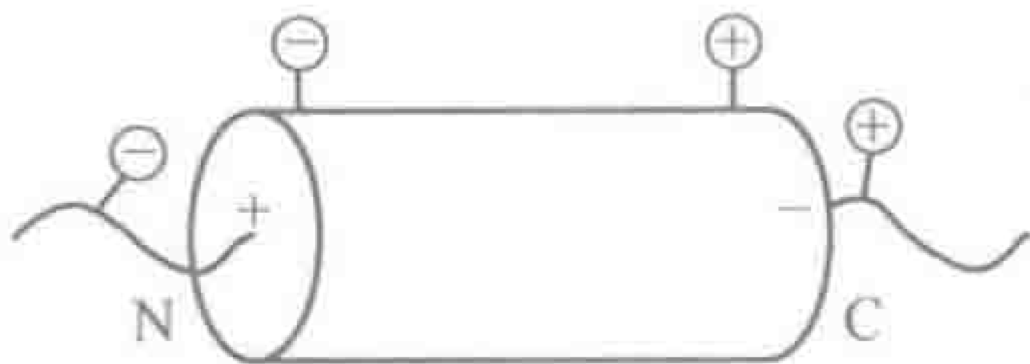


图 10-3　带电侧基团在 α 螺旋两端附近的偏好分布

至于残基在蛋白质内部或其表面的分布，其一般的趋势是，极性（**亲水的**）侧基团处于外面，在那里它们能够与极性的水相接触（“相似者相溶”）。脱离水对极性基团是不利的：会丢失氢键。这对带电基团尤其不利：离开高介电系数的媒质（水）中进入到低介电系数的媒质（蛋白质核）会导致自由能大大升高。的确，蛋白质内部实际上完全没有离子化的基团（而几乎所有特例都与金属离子的配体键有关，或者与活性中心有关：蛋白质正是为它们而制造出来的……）。

相反，大部分疏水侧基团处于蛋白质内部，正是它们在这里形成了疏水核（又一次“相似者相溶”！）。我们已经讲到，基团的疏水性越大，其非极性表面越大：正是非极性表面需要与水隔离。对于纯非极性基团，疏水效应正比于它们的表面积，而对混有极性混合物的基团则正比于扣除这些混合物的表面积后的表面积。

疏水基团的粘连是形成蛋白质球的主要驱动力。是主要的，但不是唯一的，还有在二级结构中氢键的形成（前面已经讲过），以及蛋白质内部紧密的准晶体拼装（将在后面讨论）。

为了形成蛋白质疏水核，肽链应该带着已经饱和的氢键才能进入核中，否则它的极性肽基团需要与水脱离，而氢键的断裂很费能量。因此参与到疏水核中的肽链都已经形成（或在此过程中形成）二级结构，且主链肽基团的氢键已完全饱和。然而在这种情形下在疏水核中应该只含钟爱二级结构的疏水残基，而二级结构所含的极性残基应该留在核外，因此疏水表面与亲水表面在 α 螺旋片段和 β 结构片段上分开；蛋白质链中相应基团的特定交替对于它们的形成是必需的（图 10-4）。

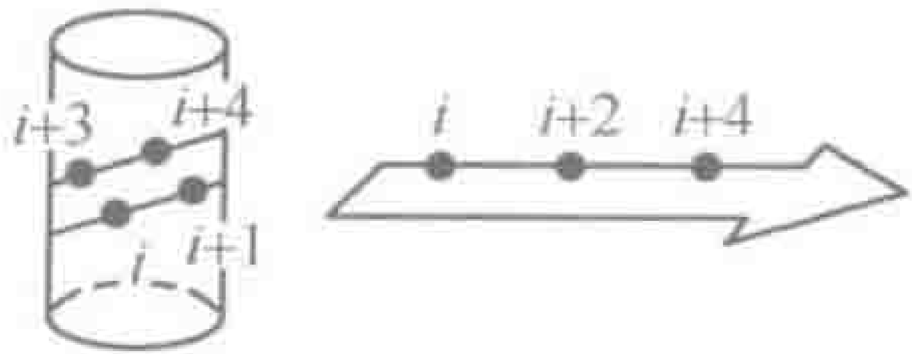


图 10-4　能够在 α 螺旋和在 β 结构片段上形成一致的疏水表面的侧基团（如果它们全都是非极性的）。链中极性基团的类似组合可导致在 α 螺旋和 β 串的反面形成亲水区

所有我们现在讲到的规律，不仅用于人造蛋白的结构设计，而且用于预测（根据氨基酸序列）蛋白质二级结构，还用于预测肽链上深埋于蛋白质内部的片段，或相反的，位于蛋白质表面的片段。对此我们将在后面讨论。

■ 最后，还对带电的（或者更准确些，可离子化的）侧链基团讲几句。pH 的升高，即 H^+浓度的降低，总是使基团“更相反地带电”：中性基团获得负电荷，而带正电的基团放电，如图 10-5 所示。不同的基团上从不带电状态到带电状态、或从带电状态到不带电状态的转变发生在不同的 pH 条件下，然而转变宽度在这里总是相同的：约为 2 个 pH 单位（在此区间内带电形式与不带电形式之比从 10∶1 变化到 1∶10）。

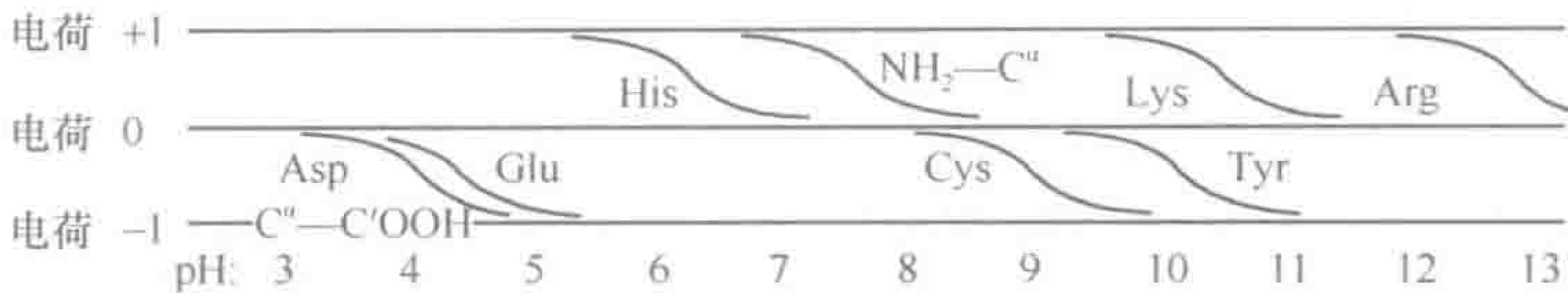


图 10-5 在水环境中不同 pH 下，可离子化侧链基团以及多肽链的 N 端（NH_2—C^α）和 C 端（C^α—C′OOH）的带电状态。对应于基团“半带电”状态的 pH，就是该基团离子化的 pK（参见表 10-2）。可离子化基团的带电状态和不带电状态的概率比对带正电的基团为 $10^{(pK-pH)}$ ∶1，对带负电的基团为 $10^{-(pK-pH)}$ ∶1

■ 需要特别注意在 pH 接近于 7（细胞中的蛋白质生活的典型值）时从不带电状态转变到带电状态的基团：正是这些容易重新带电的基团［尤其是组氨酸（His）］常常被用在蛋白质的活性中心里。

如前所述，可离子化的基团容易以放电后的状态包埋到非极性环境中（如蛋白质或细胞膜的深处）。确实，根据我们已经做过的估算，离子包埋的代价很高，每摩尔几十千卡的自由能。而可离子化基团的放电又值多少自由能呢？如果眼前有图 10-5，就不难估算。在水环境中，带正电的基团的不带电状态概率为 $W_0=1/[1+10^{(pK-pH)}]$，而带负电的基团为 $W_0=1/[1+10^{-(pK-pH)}]$。这种不带电状态的自由能为 $F_0=-kT\ln W_0$，即对于带正电的基团，若 pH>pK 则 $F_0\approx 0$，若 pH<pK 则 $F_0\approx +2.3kT(pK-pH)$。而对于带负电的基团，若 pH<pK，则 $F_0\approx 0$，若 pH>pK，则 $F_0\approx -2.3kT(pK-pH)$。因此，对图 10-5 中所示的所有可离子化基团，在“中性”pH≈7 时的“放电自由能”F_0 不超过每摩尔几千卡。

蛋白质的空间结构

第 11 讲

纤维蛋白，其功能和周期性一级与二级结构。α 角蛋白，β 丝蛋白，胶原。长 α 螺旋和宽 β 片的拼装。形成基质的蛋白质；弹性蛋白。蛋白质遗传缺陷与疾病。淀粉样纤维。

现在，我们已经了解了多肽二级结构的性质和氨基酸残基的性质，终于开始讨论蛋白质了。

前面已经讲过，按照“生活条件”、使蛋白质结构稳定的相互作用，以及基本结构类型，蛋白质可以分为三类：①纤维蛋白；②膜蛋白；③水溶性球状蛋白。

首先来看纤维蛋白。

纤维蛋白的作用主要是结构性的。它们构成微丝、微管、毛发、丝绸和其他保护性覆盖物；它们加固细胞膜并维持细胞和组织的结构。纤维蛋白通常很大。其中包括已知最大的蛋白质：肌联蛋白（titin），它是由大约 30 000 个氨基酸残基构成的。

纤维蛋白经常形成巨大的聚集体，它们的空间结构具有高度的规律性，基本上是由极大的二级结构蛋白质组装而成的，很大程度上是由各个多肽链之间的相互作用维系的。纤维蛋白的一级结构也非常有规律，有周期性，因此而形成广泛的规则二级结构。

典型的有代表性的纤维蛋白有如下几种。

■ （1）β 结构的丝蛋白。正如我们已知的，β 片的周期性在于残基的朝向交替地“向上”和“向下”（图 11-1）。

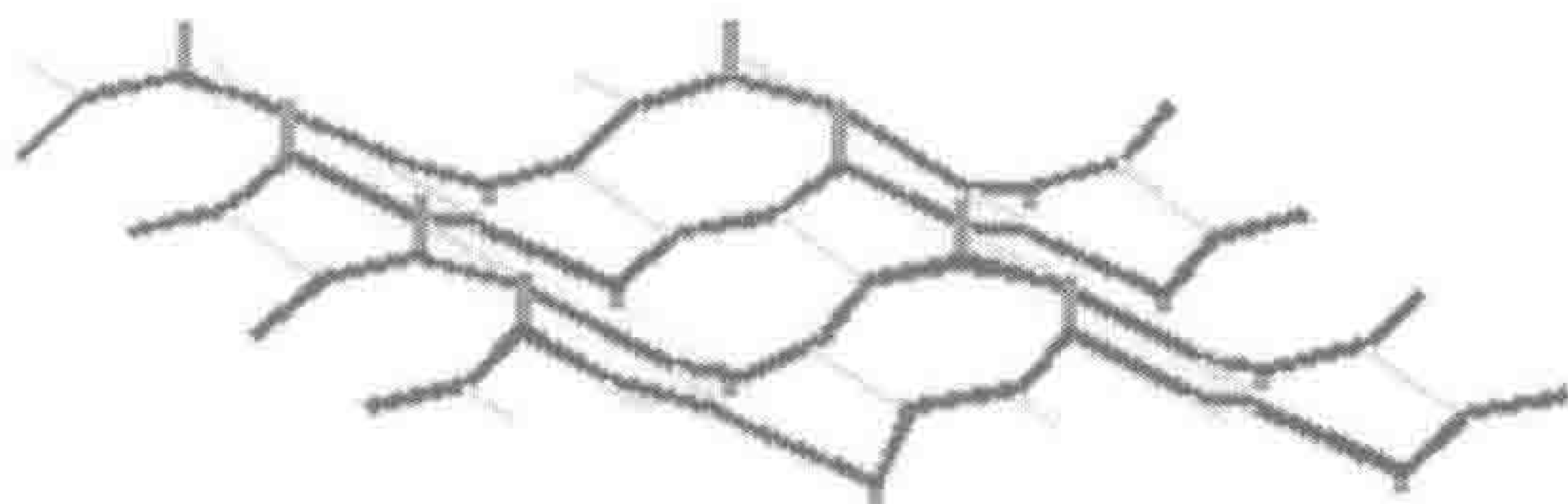
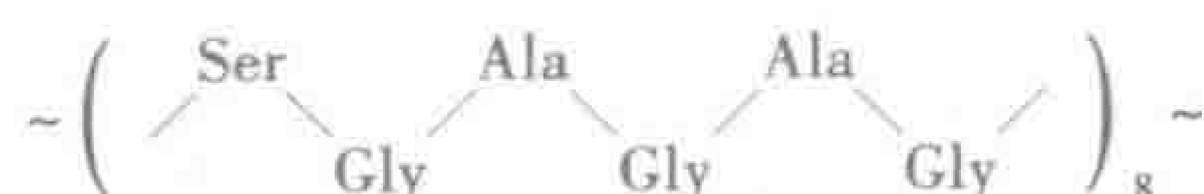


图 11-1 β 结构片。粗线强调显示它的褶皱和周期性。细灰线表示形成 β 片的肽链之间的氢键

相应地,在丝蛋白中,一级结构的基本模段(motif)看上去就是八倍体小块的重复,每一个小块中都出现有小残基甘氨酸(Gly)和较大残基的交替,例如:



这种六残基的八倍体小块,中间以某些不那么规则的链小块相连接,重复大约 50 次。

丝蛋白的反平行(如图 11-1 中的)β 层是由相互“脸朝脸,背靠背”的方式组装而成的:两层甘氨酸(Gly)(平面之间的距离为 3.5Å)—两层丙氨酸(Ala)/丝氨酸(Ser)(平面之间的距离为 5.7Å;X 射线很清楚地观察到)—两层甘氨酸……如此继续。

在丝纤维中,这些由很多 β 层组成的准晶体浸入在不那么规则的基质中。基质是由丝蛋白的不规则部分和丝胶(一种特殊的无规则基质蛋白,由二硫键结合成为巨大的聚集体)形成的。

■ (2) 由长卷曲螺旋(coiled coil)组装而成的 α 结构纤维蛋白(图 11-2)。在 α 角蛋白中,在所有具有全然不同的一级结构的中间丝的蛋白质中,在原肌球蛋白(tropomyosin)中这种螺旋环绕着所有蛋白质链,在肌球蛋白(myosin)链的大部分中也形成这种类型的纤维。这种结构在丝中也存在,但不是在前面讲到的普通的桑叶蚕的丝中,而是在蜜蜂和蚂蚁的丝中。

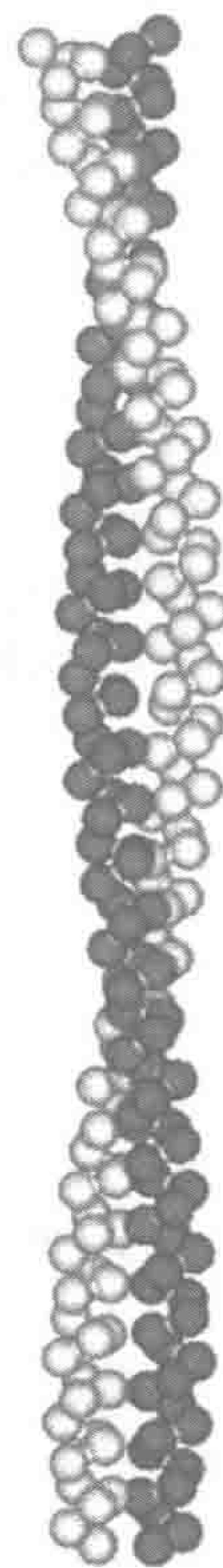


图 11-2 缠绕的右旋 α 螺旋。在复合体中它们相互平行,并相互围绕略微拧紧,每一个都形成左旋的周期为 140Å 的超螺旋。接触的氨基酸残基在链上占据重复的位置 **a** 和 **d**(图 11-3,图 11-4)

如果黏联在一起,则独立的链会形成超螺旋(图 11-2,图 11-3)。

在显示于图 11-2 中的下一级结构水平,超螺旋经常(但不总是,如并不在原肌球蛋白中)互相黏联并形成纤维。

组成超螺旋的 α 螺旋通常是平行的,它们(在不同的蛋白质中)以 2、3 或 4 条进行缠绕。正如我们已知,α 螺旋中的周期为每圈 3.6 个残基。在缠绕的螺旋中,周期为两个 α 螺旋圈 7 个残基,即每圈 3.5 个残基(图 11-3,图 11-4)

因此,典型的一级结构具有同样的周期性(7 个残基),外表如图 11-4 所示。在该图中粗体字母对应于疏水的(油脂的,黏性的)氨基酸,其他字母对应于亲水氨基酸。

有趣的是,处于“过渡”位置 *e* 和 *g* 的残基的疏水性稍稍增大,会使二联体缠绕的螺旋[图 11-3(a)]变为三联体[图 11-3(b)],如果更大会变成四联体,如此等等。

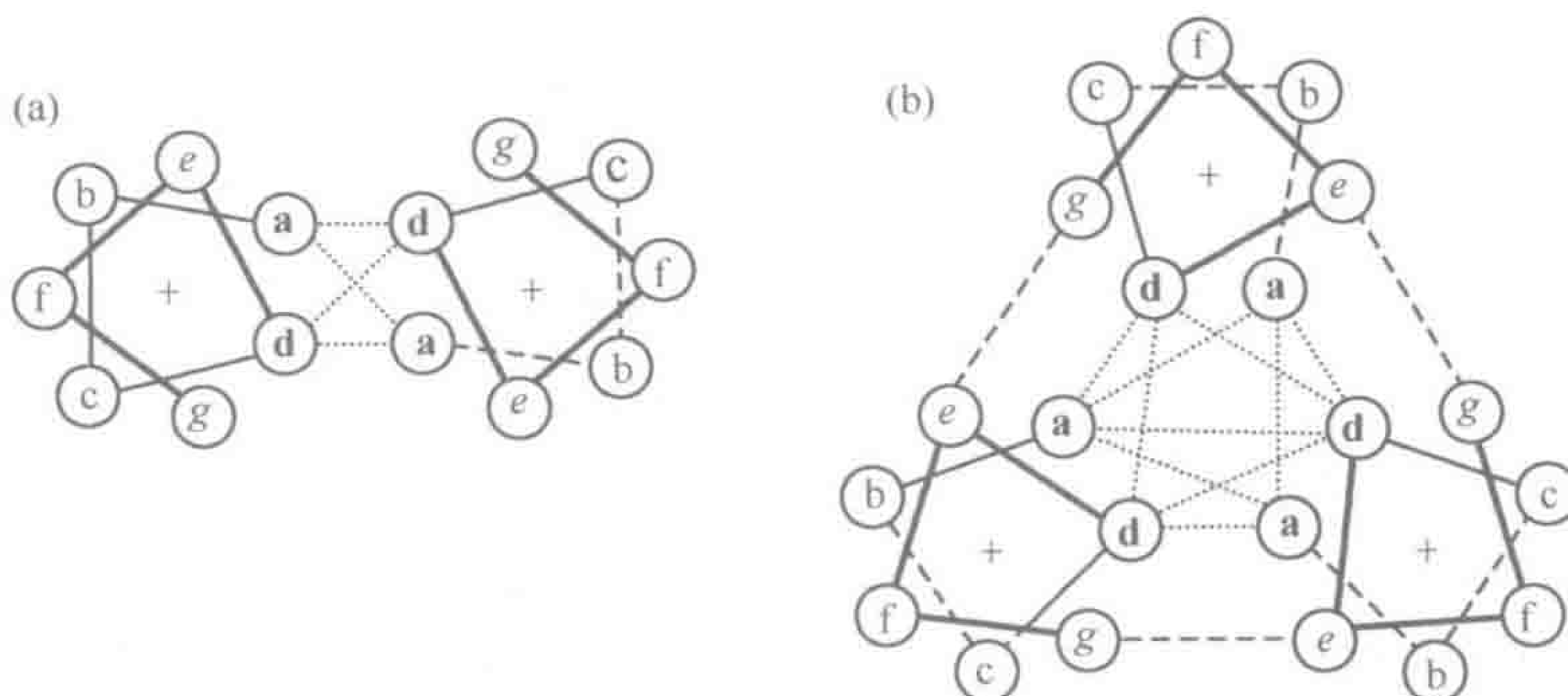


图 11-3　二联体(a)和三联体(b)超螺旋中的 α 螺旋的相互作用(螺旋顶部观)。在二联体超螺旋中螺旋只有残基 **a** 和 **d** 直接接触,而在三联体中,还在残基 *e* 和 *g* 之间接触(虽然更弱)。图片取自文献[6],稍有修改

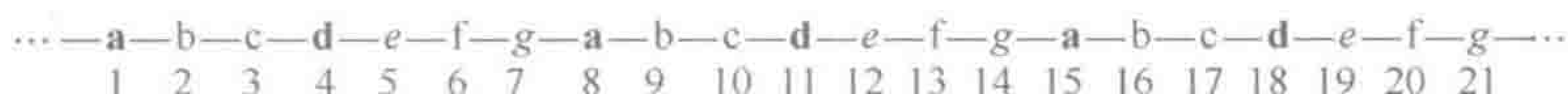


图 11-4

同样有趣的是,湿润的由 α 螺旋组装而成的纤维在机械张力的作用下会转变为 β 结构形式,而在移除张力并降低湿度时,又可恢复为 α 螺旋结构。

■　让我们更详细地看一看,螺旋是如何彼此粘连的。在 α 螺旋[图 11-5(a)]上有几条螺旋线,在空间上靠近的侧链基团构成的"脊骨"。一种脊骨具有 1–4–7–…的周期形式,而它们位于接触区域内的部分,是由残基对 $\mathbf{a}_1$–$\mathbf{d}_4$,$\mathbf{a}_8$–$\mathbf{d}_{11}$,…组成的[图 11-5(b)]。另一种脊骨具有 0–4–8–12–…的周期形式。它们位于接触区域内的脊骨部分,是由残基对 $\mathbf{d}_4$–$\mathbf{a}_8$,$\mathbf{d}_{11}$–$\mathbf{a}_{15}$,…组成的[图 11-5(c)]。

第一种脊骨,"$i,i+4$"(如果以侧链基团形成的突出部标示它们)型脊骨与螺旋轴的夹角约为−25°,第二种为"$i,i+3$"型脊骨约为+45°(图 11-5)。如果绕垂直轴翻转,把一个螺旋的表面置于另一个之上[图 11-6(a)],然后绕平面的垂直轴转动+20°,则"1–4–7"型脊骨落在另一个螺旋的"0–4–8"型脊骨之间,就会保证它们紧密接触[图 11-6(d)]。此时一个螺旋的 **a** 基团落在另一个的 **d** 基团之间,而接触区域在两个螺旋(此时它们并未形成超螺旋)表面形成略微绞合的螺旋条带。而当这些螺旋形成超螺旋(图 11-2)并围绕公共轴缠绕时,接触区域出现在略微扭转的束的中心。

这并不是螺旋紧密接触的唯一方式(将在讲述球蛋白时认识其他方式),但却是对纤维蛋白典型的极长螺旋有利的唯一方式。这是由 Francis Crick 在 1953 年(也就是他与 James Watson 一起预言 DNA 的双螺旋结构的时候)预言的。

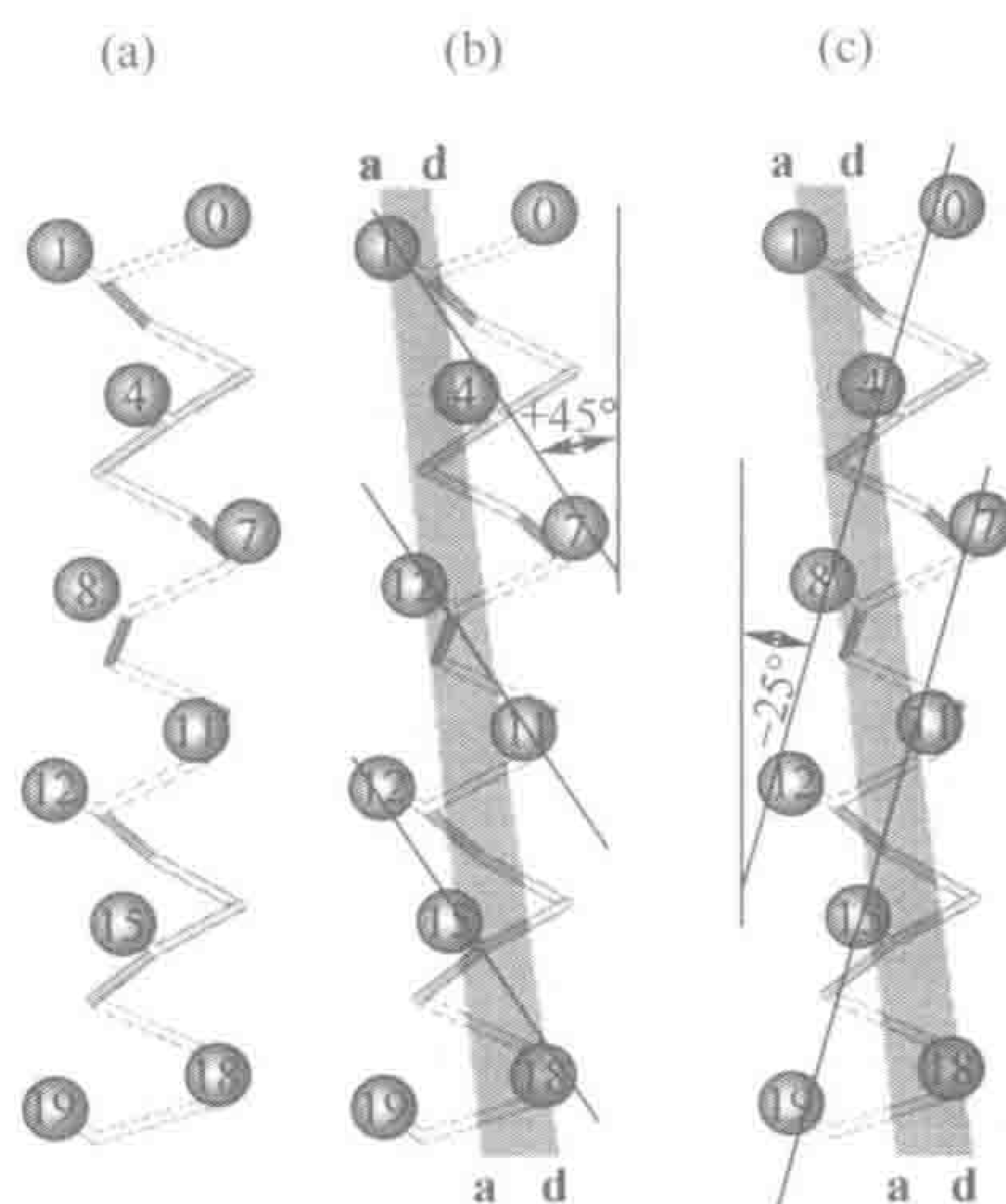


图 11-5　α 螺旋(主链和 C^{β} 原子)和在其表面上靠近的侧链基团构成的两种脊骨类型(细线)。(b)$i,i+3$ 型(或简称“+3”)脊骨;(c)$i,i+4$ 型(或简称“+4”)脊骨。接触表面以条带显示;其边缘标示了残基列 **a** 和 **d**。图中还标出了 $i,i+3$ 型脊骨和 $i,i+4$ 型脊骨相对于螺旋轴的典型倾斜角(图中的角度显示得较小,因为典型的脊骨穿过大量的侧链基团,而在图中是穿过 C^{β} 原子的中心)。图片取自文献[5],略有修改

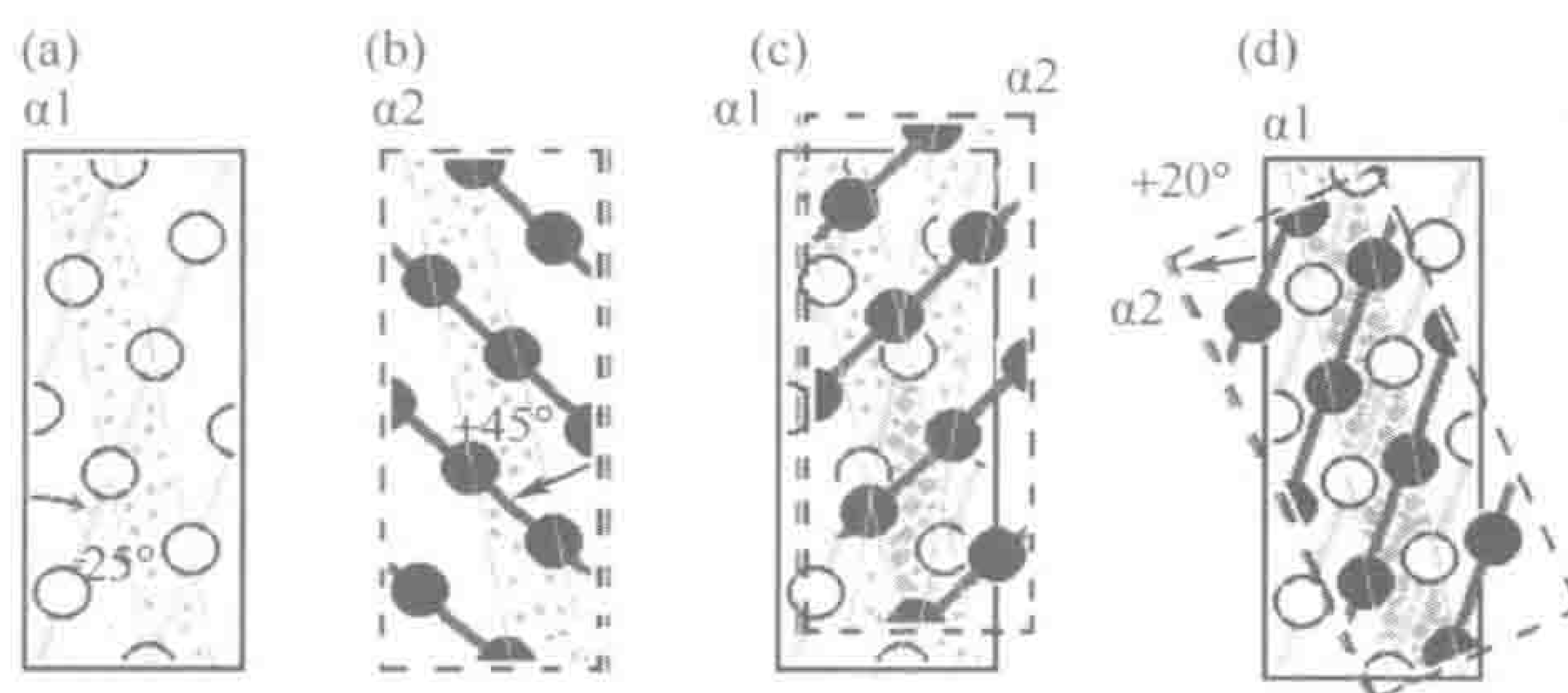


图 11-6　在螺旋接触时的侧链基团脊骨的紧凑拼装要求一个螺旋相对于另一个转动+20°。我们透过一个螺旋(逆向沿 α2 轴)观察接触区域。“下面的”螺旋 α1 的残基以浅色圆圈显示,上面的(α2)以深色圆圈显示。细线表示接触残基 **a** 和 **d**。图片取自 Chothia, et al. , 1977. Proc Natl Acad Sci USA.74:4130-4134,略有修改

■　(3) 胶原蛋白(collagen,希腊语原意为“形成胶的”)。它是主要的结构性蛋白质,占到脊椎动物蛋白质质量的大约 1/4。它形成坚固的不溶于水的纤维。胶原蛋白分子由特殊的三联体超螺旋构成,每个超螺旋由 3 条多肽链组装而成(图 11-7,参见彩色立体图对 D-11)。在这种方式下,在每条多肽链的内部和三联体螺旋丝的内部,都没有氢键,只在丝之间存在有氢键。

每条胶原蛋白链中的所有残基的构象与聚脯氨酸[更准确一点:poly(Pro)II]螺旋的构象很相像;然而,与 poly(Pro)II 不同的是,胶原蛋白螺旋具有链间氢键。

它是左旋的螺旋，周期等于3。相应地，胶原蛋白一级结构的基本模段是三残基组(Gly-Pro-Pro)$_n$ 的多次重复，或更确切些，是(Gly-某种残基-Pro)$_n$，而且甘氨酸(Gly)在每个三残基组中对于形成氢键是必不可少的：它含有NH基团(这与Pro不同)，并且没有侧链基团。形成氢键的甘氨酸处于在紧密的胶原蛋白超螺旋的中心，这里任何侧链基团都是多余的。

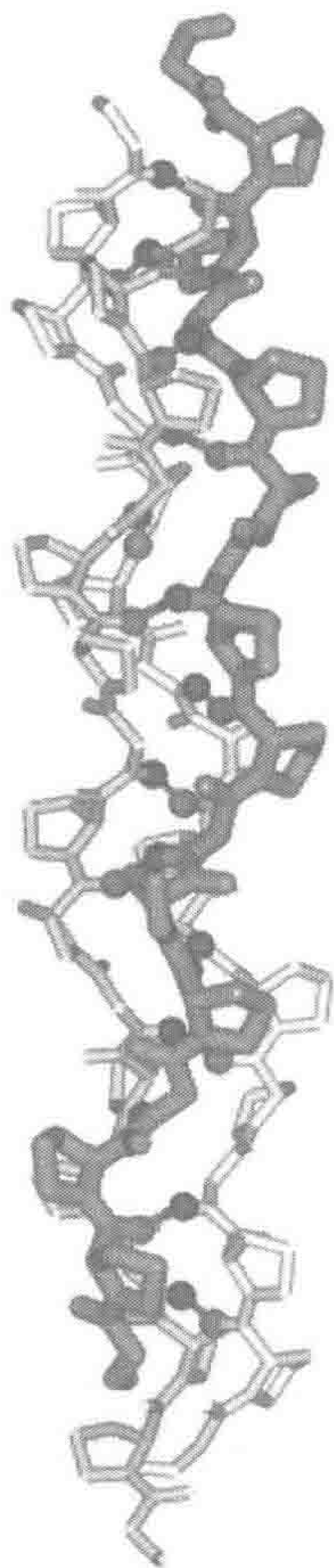


图11-7 序列(Gly-Pro-Pro)$_n$ 的胶原蛋白三联体超螺旋模型。各链分别以不同颜色显示。Gly的NH基团的H原子与三残基组Gly-Pro-Pro中的第一个Pro的CO基团的O原子结合而成的氢键以小黑哑铃表示。链“1”的Gly与链“2”结合成键，而Pro与链“3”结合成键，如此等等。胶原蛋白的每一条链都与其他两条链缠绕，形成**右旋**的超螺旋。被称为“超螺旋”是因为在更小的尺度，在单个残基构象的尺度，胶原蛋白链已经形成了poly(Pro)II型的螺旋(而且这种“微螺旋”是**左旋的**)；这可以根据Pro的环的方向而看出来。胶原蛋白分子直径约15Å，长度约3000Å

有趣的是，编码胶原蛋白链的外显子(exon)，总是以甘氨酸开始，而且编码数目总是3的倍数。我希望，你们记得，真核生物的基因包含编码小块蛋白质的外显子和从mRNA剪切出来的内含子(intron)(因此不编码蛋白质)。

在更高一级的结构水平，胶原蛋白超螺旋也彼此黏联，形成胶原蛋白纤维。

胶原蛋白的生物合成、随后的修饰和胶原蛋白纤维的成熟结构的形成已经被研究得很透彻了(图11-8)。

为了自组织形成胶原蛋白，需要有**前**胶原(procollagen)，它不仅含有胶原丝，还含有球状头和尾巴。缺少头和尾的前胶原丝可以在体外环境的各种不同状态下自发形成，也能折叠成三联体螺旋，但却是“不正确的”，缺少天然胶原蛋白固有的三联体螺旋异质性(天然胶原蛋白包含两种丝)，也缺少固有的配准性(即丝间的相对偏移不正确)，以及缺乏其他性质。因此胶原蛋白**不能**在体外环境自组织形成正确的空间结构，这与蚕丝相似，而与前面讨论过的α结构超螺旋不同，特别是与后面要讲到的球蛋白也不相同。

在温度升高时胶原蛋白螺旋会熔化(这样就形成了凝胶)。胶原蛋白的熔化温度与脯氨酸(Pro)和羟基脯氨酸的含

量密切相关(这些硬性残基越多,熔化温度自然就越高),而且通常比宿主动物的体温只高出几度。请注意这个事实,几讲之后我们还会回来讨论它。

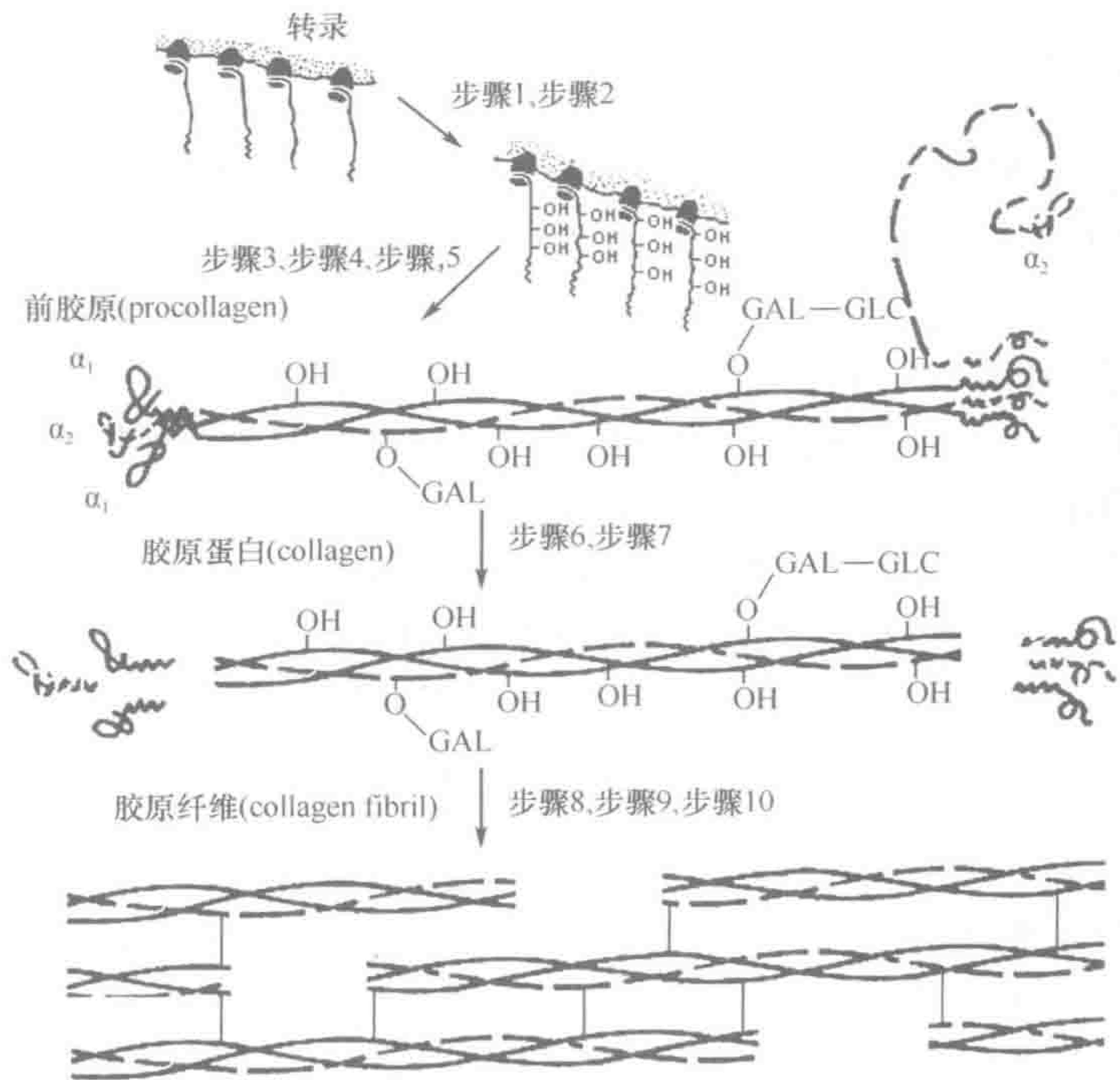


图 11-8　胶原蛋白在体内的形成。步骤 1:pro-α_1 链和 pro-α_2 链(每个链中有 1300 个残基)按 2∶2 比例进行生物合成。步骤 2:某些位于 Gly 前的 Pro 和 Lys 残基的酶促共翻译羟基化。步骤 3:糖(GLC-GAL)缔合到一些羟基化的残基上。步骤 4:由两个 pro-α_1 链和两个 pro-α_2 链的 C 端小球形成四聚体;其中一个 pro-α_2 降解;形成前胶原(pro-α_1-α_2 三聚体),并在其球形末端上形成 S—S 键。步骤 5:在前胶原中部形成三倍体螺旋。步骤 6:前胶原分泌到细胞外空间。步骤 7:球状部分被剪除;步骤 8~10:由三联体超螺旋自发形成纤维,氨基酸残基的最终修饰,和前胶原链上的被修饰的残基之间形成共价交叉偶合(在特异性酶的作用下)。图片取自文献[3],有改编

■　胶原蛋白的折叠特别让人感兴趣,原因在于有多种遗传性疾病与影响胶原蛋白折叠的突变有关。其中研究得最多的疾病是"脆骨病"(osteogenosis imperfecta)。这种病症的最常见的原因是有一个核苷替换,导致了 Gly 被替换成另一种氨基酸。这破坏了$(\mathrm{Gly}\text{-}X\text{-}Y)_n$型的重复序列,该序列是胶原蛋白的三螺旋特征结构的基础。看来,至少该突变的后果之一是胶原蛋白折叠变慢:慢得能允许该蛋白质发生不正常的翻译后修饰。为了克服"完整"胶原蛋白研究上相关的困难,与疾病相关的突变效果是在简化的周期性多肽上进行研究的。这允许进行详细的生物物理学研究,其中包括"实时"NMR 实验,以探索在单个残基水平上的折叠步骤。顺便说一句,这一示例清楚地说明了正在寻找中

的蛋白质错误折叠与人类疾病之间的联系，也说明了在对这些疾病的分子基础的研究中恰当地应用结构学方法的威力。

■ 在结尾，我想强调一下，纤维蛋白在结构上常常是很简单的，原因在于它的一级结构的周期性，以及二级结构也具有周期性。

■ 然而，这里再做一点补充也是合适的。

有些蛋白质能形成巨大的、内部没有任何确定结构的聚集体，常常也被归类于纤维蛋白。这些蛋白质形成以某种化学性连接相缝合的弹性基质，其他一些更具结构性的蛋白质浸于其中。

弹性蛋白(elastin)是典型的基质蛋白。它在动脉壁、肺壁、韧带等的构成中发挥着重要的作用。它的长链相当地疏水，且包含几种类型的短残基重复。弹性蛋白的每条链形成无序的线团，但这些链被由酶修饰的赖氨酸(Lys)缝合成网(每四条在一个节点缝合)。结果是形成了某种类似橡胶的东西。我忍不住提及，在弹性蛋白中赖氨酸修饰酶的工作被破坏时，会导致韧带和血管的弹性丧失，有时甚至导致大动脉破裂。

■ 最后。最近观察到，许多"正常的"水溶性球蛋白(至今还没有关于膜蛋白的数据)能够形成所谓的淀粉样纤维。其中有溶菌酶(lysozyme)、肌红蛋白(myoglobin)及其他，尤其是它们的几个突变体。但是，首先来看著名的普利昂蛋白(prion)。关于淀粉样纤维很久以前就知道，它们不与蛋白质相互结合——直到人们发现，普利昂蛋白，这些"感染性蛋白"，在它们所损害的细胞中形成了淀粉样纤维。后来查明，其他最普通的蛋白质也能形成与此相似的淀粉样蛋白，而在此背景上，普利昂蛋白看来只是由于其感染性而更为显著。在所研究的淀粉样纤维中，蛋白质链组装形成双层或多层由平行 β 结构构成的夹层。β 串垂直于纤维轴，而它们之间的氢键沿该轴走向(图 11-9)。β 夹层紧凑拼装成管状或丝状(其形状依赖于蛋白质及外部条件)的原纤维，然后堆码成长长的淀粉样丝。看来，沉浸在 β 结构中的并非整个蛋白质链，而只是其中一部分[而其他部分维持球状(或解折叠?)并处于淀粉样原纤维骨架之外]。看上去，在淀粉样丝中，只有蛋白质的那些参与到形成原纤维的 β 夹层的部分具有唯一的空间结构。已知，淀粉样纤维具有缓慢(进行时常常带有"计程器"，即在开始增长之前有很大的延迟)且不可逆的 β 结构聚集。有充分的理由认为，对机体的损害不只是成熟而较大的淀粉样纤维产生的，而且是在它们形成微小的活性低聚"晶核"的过程中就产生了。无论如何，形成淀粉样纤维的过程和/或结果会导致潜伏期可能很长的严重疾病(库鲁病①、

---

①库鲁病(Kuru)是最早被研究的人类蛋白质病毒疾病，曾经仅见于巴布亚新几内亚东部高地有食用已故亲人脏器习俗的土著部落，自从这一习俗被废止后已无新发病例。库鲁病潜伏期长，4~30 年不等，起病隐匿，前驱期患者仅感头痛及关节疼痛，继之出现共济失调、震颤和不自主运动，后者包括舞蹈症、肌阵挛等，在病程晚期出现进行性加重的痴呆和神经异常。患者多在起病 3~6 个月内死亡。

狂犬病和阿尔茨海默病①等)。对这些以及其他纤维蛋白(和其他蛋白)的"构象病"的研究是当代分子生物学的"热点"之一。

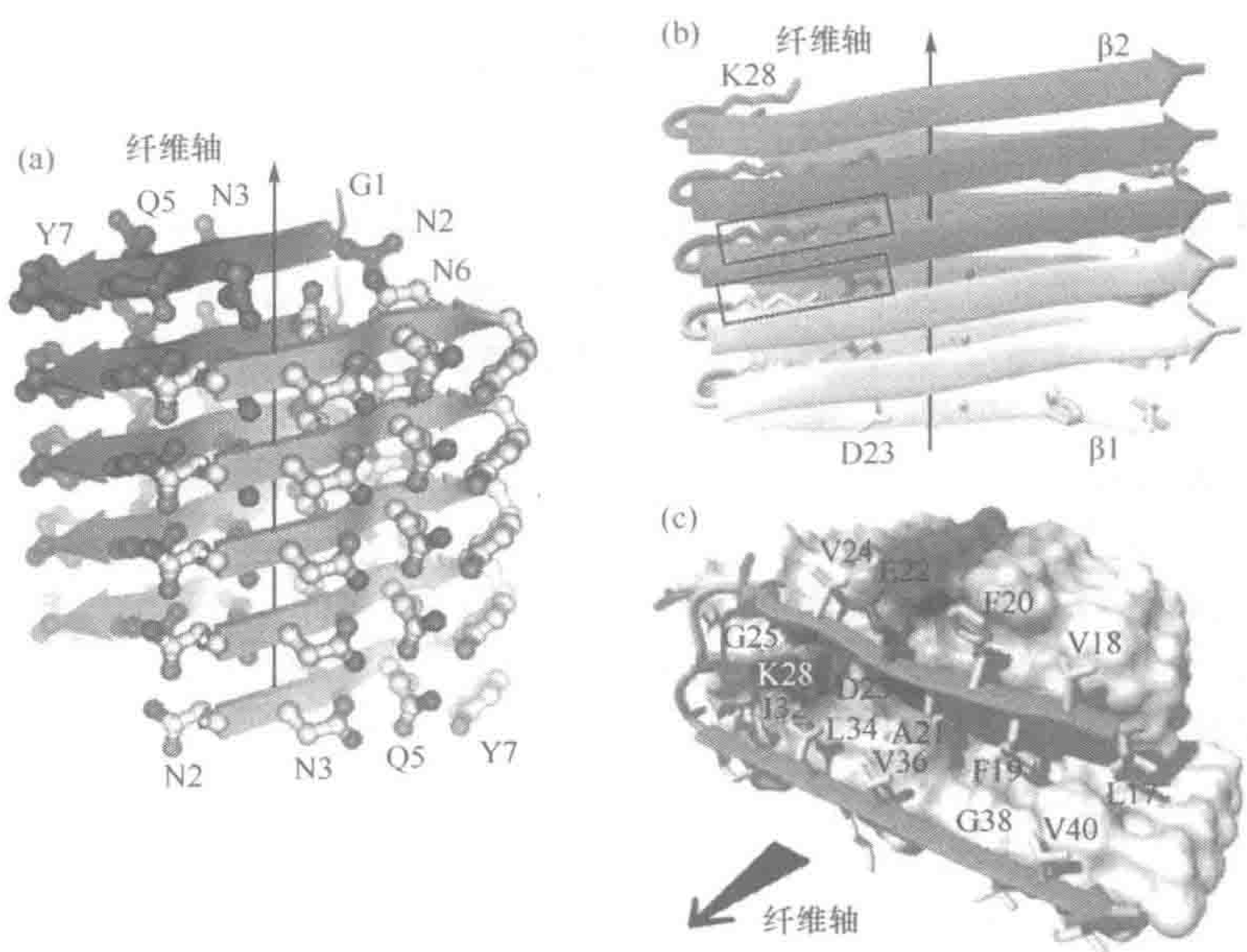


图 11-9 (a)通过 X 射线结构分析得到的淀粉样纤维小块,由酵母蛋白 Sup35 的短(7 个残基)片段构成(在酵母中引起的疾病与普利昂蛋白在哺乳动物中引起的疾病类似)。图片取自 Nelson,et al. ,2005.Nature.435:733-778,略有简化。请注意,该纤维的"疏水核"是由极化氨基酸(极性大于水)Asn(N)和 Gln(Q)构成的,它们的侧链基团在彼此之间形成了氢键。(b)和(c):比较间接地用核磁共振方法解析的淀粉样纤维从不同角度的外观。该淀粉样纤维是由取自引发阿尔茨海默病的蛋白质的 42 残基的多肽 Aβ(1-42)组成的。图片取自 Lührs,et al. 2005. Proc Natl Acad Sci USA.102:17342-17347,有简化。在多肽 Aβ(1-42) 中残基 1-17 的部分是无序的(因此在图中只能从 Leu17 开始显示),而 Aβ(1-42) 的其他部分进入了原纤维的骨架,骨架的疏水核主要是由"正常的"无序氨基酸组成的,虽然也包括了离子对 Asp23-Lys28

■ 另外,由"正常的"、未经过重组的球蛋白组装成的纤维(如肌动蛋白丝②)有时也认为是纤维蛋白。我们在这里不讨论它,在后面讲述球蛋白时再讨论。

①阿尔茨海默病(Alzheimer's disease,AD),是发生在老年期及老年前期的一种原发性退行性脑病,指的是一种持续性高级神经功能活动障碍,即在没有意识障碍的状态下,记忆、思维、分析判断、视空间辨认、情绪等方面的障碍。其特征性病理变化为大脑皮层萎缩,并伴有 β 淀粉样蛋白沉积,神经原纤维缠结,大量记忆性神经元数目减少,以及老年斑的形成。目前尚无特效治疗或逆转疾病进展的治疗药物。

②关于肌动蛋白丝,请参见第 25 讲。

第12讲

膜蛋白,它们的结构与功能的特性。细菌视紫红质,受体和G-蛋白,孔蛋白。光合作用中心。膜通道的选择通透性。光合作用中心的工作方式。隧道效应的概念。光子-构象相互作用的概念。

现在让我们来讨论膜蛋白。它们(我指的是它们的跨膜部分)的构造同样是相当简单的,几乎与纤维蛋白一样。

细胞膜构成细胞外表面,而它的内部是不同的封闭体积,或者所谓的腔室。膜是由油脂(脂类)和蛋白质构成的(图12-1)。膜蛋白(它们占膜质量的一半)的特殊作用是保证各种不同的物质以及信号能够穿过它。众所周知,膜是"绝缘体"。而蛋白质(更准确些,正如我们将要看到的,是其在膜中的通道)是"导体"。这些导体是特异性的,每一种都只容许特定的分子(或者借助于蛋白质构象的微小变化而容许来自特定分子的信号)通过膜。

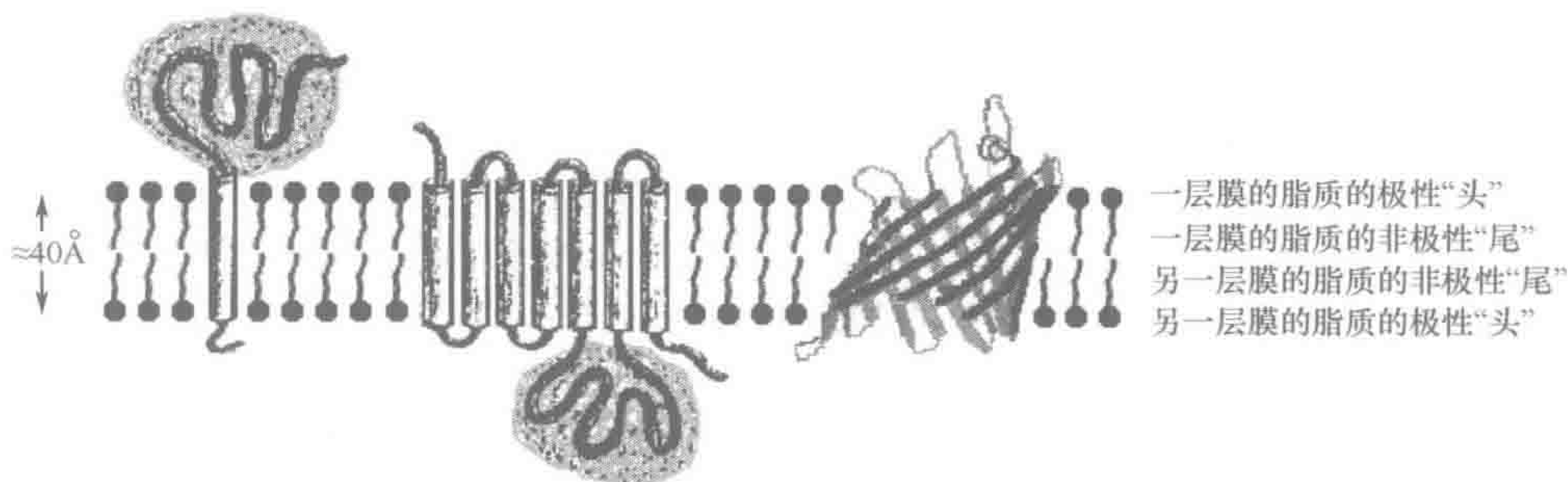


图12-1　细胞膜中的蛋白质。灰色表示膜外区域。蛋白质的膜内部分实际上并不包含不规则的链片段。真核细胞的膜蛋白从细胞向外伸出的部分,通常都是高度糖基化的,从而加强了它们的亲水性

确实,膜蛋白"住在"膜内:那里实际上没有水,因此这些蛋白的膜内部分应该是由规则的二级结构组成的(正如我们将要看到的,确实如此),而且这些部分的大小受到膜厚度的限制。

让我们来看几个例子,看膜蛋白是如何构造的。总的应该来说,我们对膜蛋白的结构知道得不多(至今总共只解析了几百个膜蛋白的结构①),因为这些蛋白质

①详表见网页:http://blanco.biomol.uci.edu/mpstruc/listAll/list.

很难溶解于水中(不得不使用去垢剂之类的[1]),而且由于它们有无规则地粘连的倾向,从而极难结晶。

■ 细菌视紫红质(bacteriorhodopsin,BR)[2](它跨膜传递光子)的结构显示于图12-2中。由于只是在相对不久前才获得适合于X射线结构分析的细菌视紫红质三维晶体,所以这个结构是在分析了大量的高精度电子显微照相的基础上构建而成的。

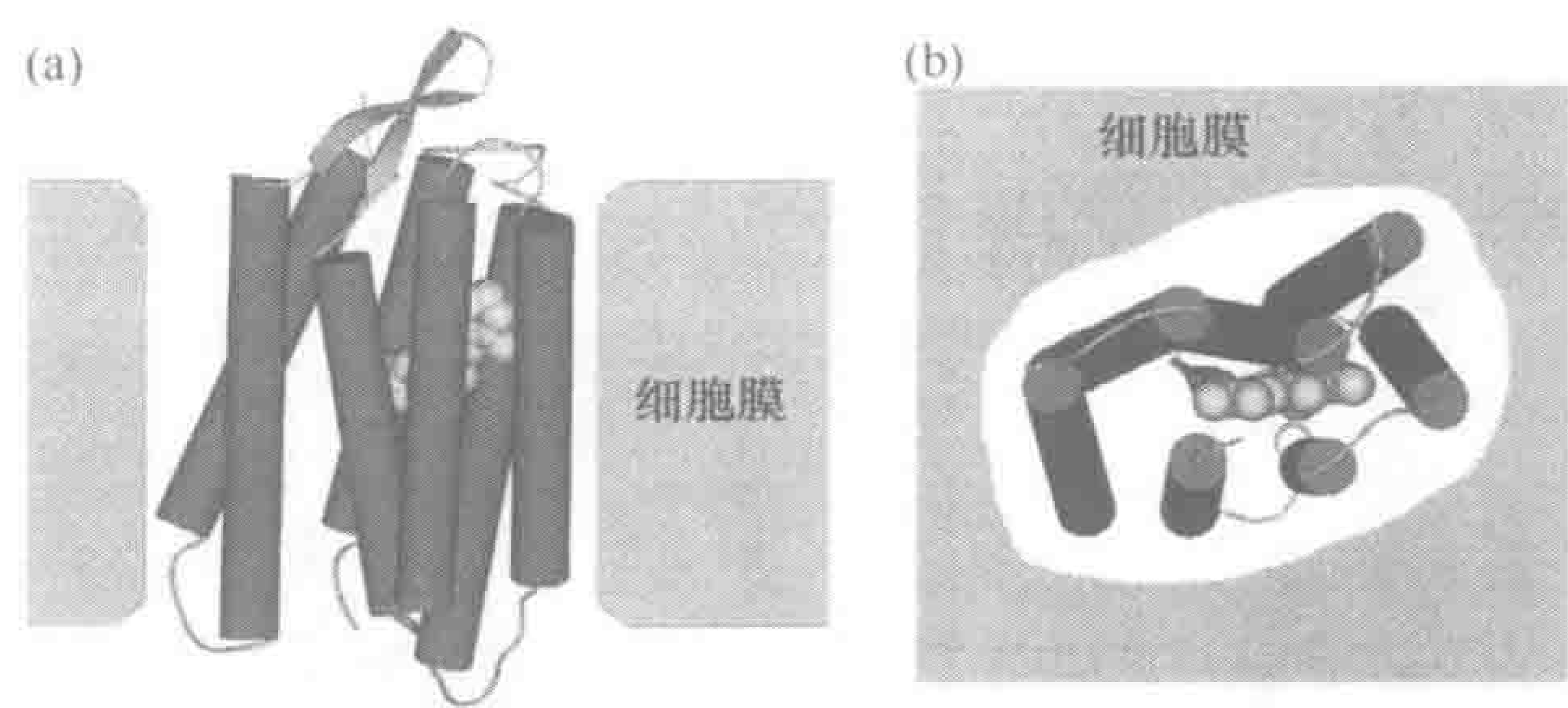


图12-2 膜中的细菌视紫红质:(a)膜剖面观;(b)膜下方观(从细胞内)。该蛋白的七个螺旋以圆柱体显示。同时还显示了与这些螺旋相邻的环线,以及黏附在视紫红质菌内部的由原子小球组装而成的视黄醛分子模型

我们看到,细菌视紫红质的跨膜部分是由7个规则的α螺旋组装而成的,它们从膜的一边穿到另一边,并形成与膜平面略微倾斜的一个螺旋束,而唯一的β发夹和所有的不规则链片段(连接螺旋的环线)从膜中穿出。

跨膜蛋白质链骨架的折叠具有高度的规则性,这是很自然的:在几乎无水的脂质"生活"环境中,每个氢键的价值都非常高,这迫使进入到膜中的任何蛋白质链都采取充分形成氢键的结构,即要么是α螺旋,要么是β结构(更确切些,是使β片不会暴露在外面的β圆筒。见下文)。

细菌视紫红质α螺旋上的疏水基团朝向"外面",指向(同样疏水的)脂质膜。极性基团(它们并不多)朝向很狭窄的通道的内部,质子就是沿该通道移动的。质子的通透性是在黏附在螺旋束内部的一种辅因子视黄醛的协助下得以实现的。它能阻塞细菌视紫红质的中心通道。

一旦吸收了光子之后,视黄醛从反式构象就转变为顺式构象。然后它弯曲(而且蛋白质体的构象也略有变化),并把质子从细菌视紫红质的七螺旋束的一面(细胞内的,图12-2中的下方)转移到另一面(朝向细胞外表面,该处自由质子比细胞

---

[1]去垢剂是同时具有亲水极性基团和疏水非极性基团的双极性分子,能够使脂膜解体释放膜蛋白,并在溶液中为去膜状态下的膜蛋白提供疏水环境,维持和保护膜蛋白的疏水跨膜结构。

[2]细菌视紫红质是盐生嗜盐菌等细菌的跨膜蛋白质,光照时它吸收光子,其色基视黄醛的光致异构化作用触发细菌视紫红质的一系列结构变化,把质子从细胞质泵到细胞外空间。

内要多)。然后视黄醛伸直并恢复原形,但已经没有质子了,直到下一个质子来到。

■ 在其他条件下,由分立的 α 螺旋跨膜多肽也能形成类似的由螺旋空心束构成的孔洞(但是通常更宽)。

顺便说一句,螺旋束常常出现在完全不同的膜蛋白中。它们不跨膜传递任何分子,而是传导信号。我指的是各种不同的蛋白质受体,尤其是激素受体(图 12-3)。这类受体有很多,我们来简短地讨论一下其中一个的工作机制。

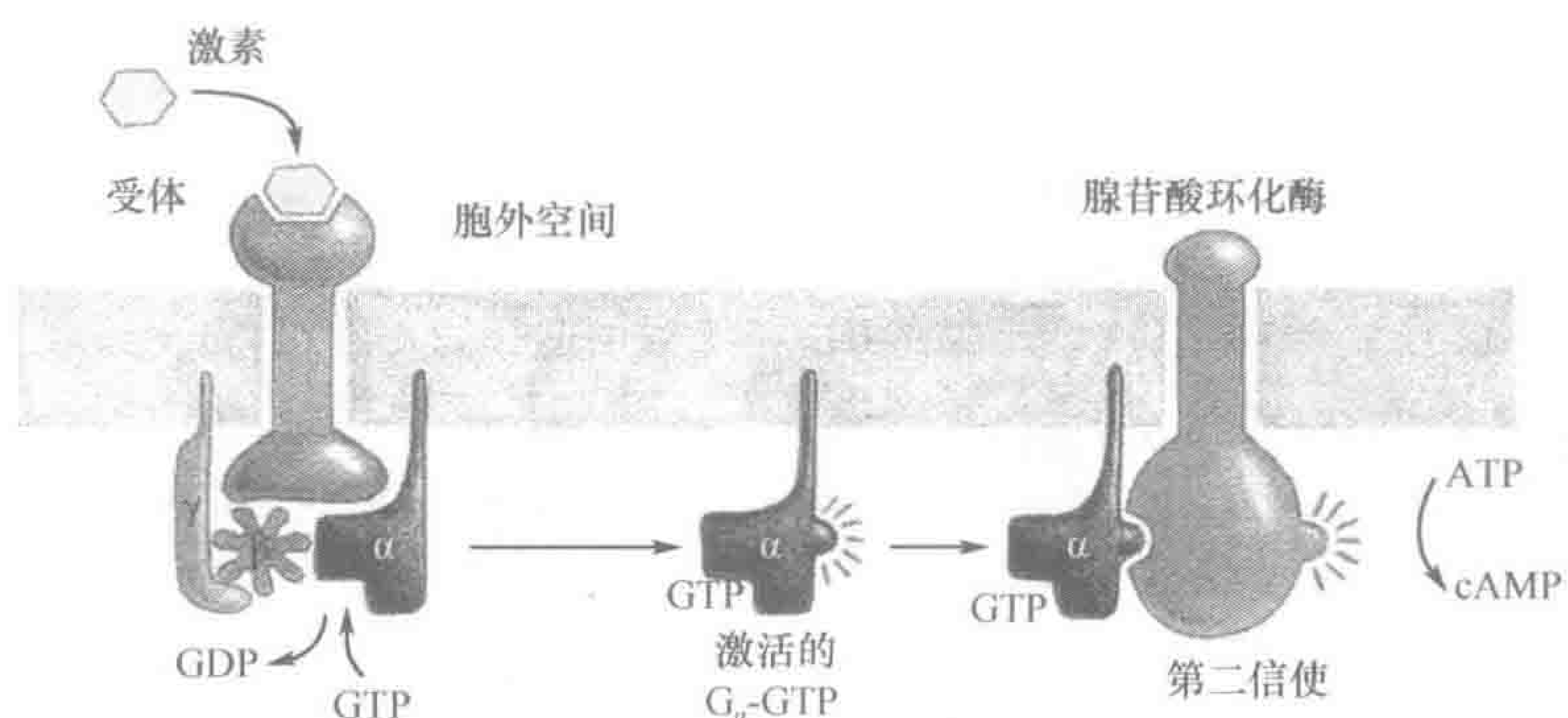


图 12-3 腺苷酸环化酶借助于 G 蛋白而激活,而 G 蛋白由与胞外空间所释放的激素相结合的跨膜受体所激活。ATP→cAMP 的转变是由胞内的第二信使所产生的。图片经作者同意取自文献[5]

在与激素结合后,这种受体的跨膜螺旋束的构象以某种方式发生改变(仍不清楚如何改变)。通过这种方式,激素到来的"信号"被传递到细胞内:受体的膜内部分的构象变化协助使与之相结合的 G 蛋白(guanidine①-binding protein)的 α 亚基释放已结合的 GDP(鸟苷二磷酸)分子,并与周围细胞溶质中的 GTP(鸟苷三磷酸)分子相结合,然后离开它的其他亚基和受体(此时可能另一个已结合了 GDP 的 G 蛋白的 α 亚基与受体靠近,同样的,G 蛋白与 GTP 结合,并释放 GDP)。

G 蛋白的 α 亚基能分解 GTP,使之变成 GDP 和磷(P),但它工作得很慢(这很重要!)。于是,它与 GTP 一起沿着膜(它的"尾巴"插入膜中)飘移,游近腺苷酸环化酶,并与之相结合,于是腺苷酸环化酶开始工作,把大量 ATP(三磷酸腺苷)分子转化为 cAMP(环腺苷酸)。这也启动了细胞对激素结合的生理响应。但 G 蛋白的 α 亚基对腺苷酸环化酶的作用持续时间并不长;最后 α 亚基把自己的 GTP 转变成 GDP,作用结束,且 α 亚基(带有 GDP)回到自己的受体上,如果受体仍与激素相结合,则重复这一循环;未结合,则结束。

结果是,一个激素分子的信号被多次放大,但只持续有限长的时间。所有 G 蛋

①胍(guanidine)是主要在甜菜汁、野豌豆苗或胚鸡中发现的一种极容易潮解的晶体碱,分子式为 $NH=C(NH_2)_2$,它由鸟嘌呤氧化生成。

白(不仅只是那些能激活腺苷酸环化酶的 G 蛋白)的专一性都归结于作用时间的有限性:它们在 GTP 尚未降解的时间段内工作(因此慢速的 GTP 降解起到了定时器的作用)。

■ 再来看另一个跨膜蛋白:孔蛋白(porin)(图 12-4)。它同样是有高度规则性的,并具有由 β 结构组装成的宽圆筒的形状。请注意,此处的 β 片形成封闭的 β 圆筒,这可以避免在平面 β 片边缘处很常见的“未结合的”氢键。在这个圆筒中有 16 个很长的 β 片段,通过其中心的孔洞的直径相当大,约 15Å。参与形成 β 串的极性残基的侧链基团朝向洞壁,而在 β 串中与之交替的非极性残基把自己的侧链基团朝向膜。

孔蛋白负责转运极性分子,不过选择性并不很强。

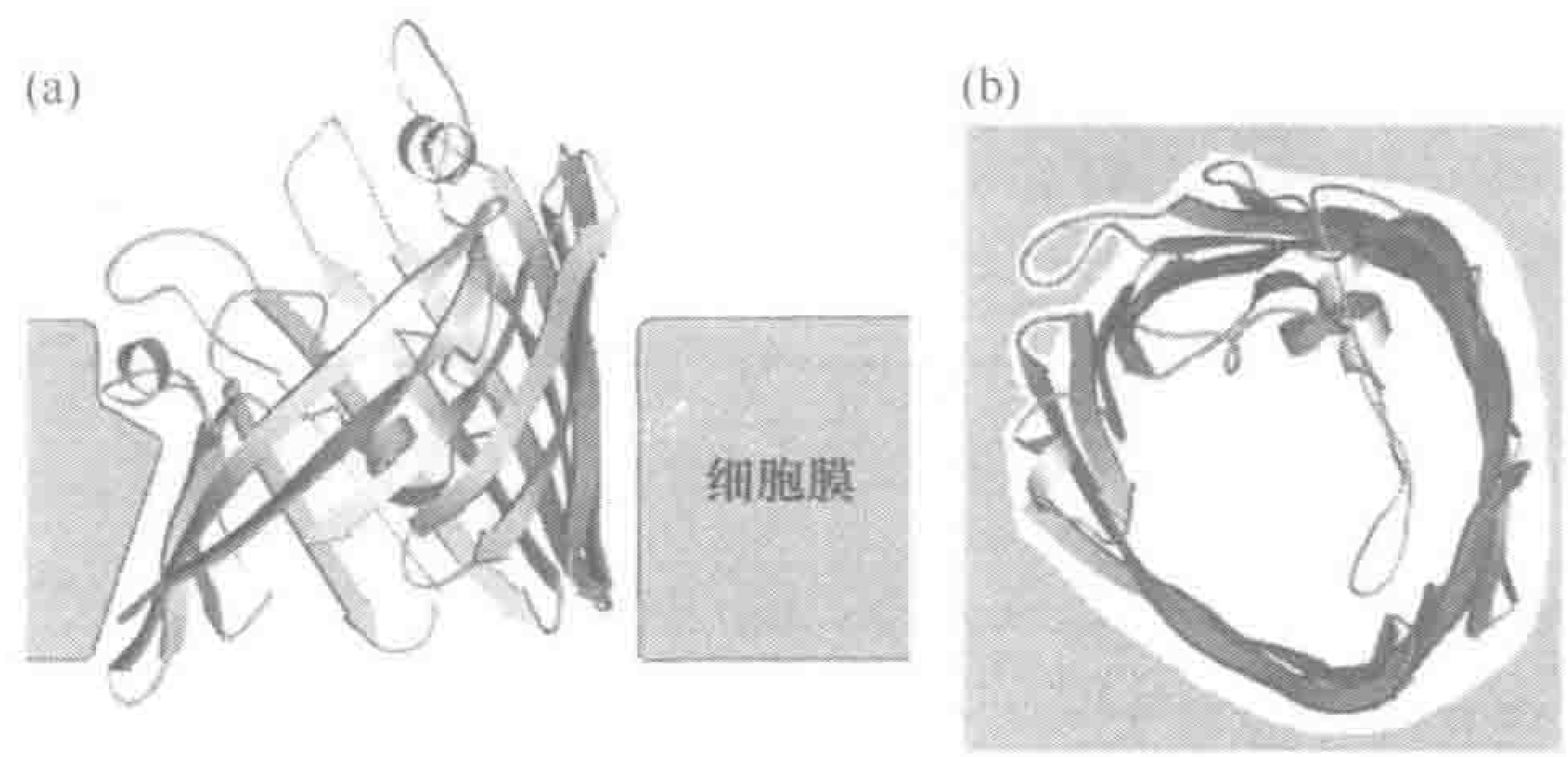


图 12-4 孔蛋白:(a)膜剖面观(膜中的脂质只是示意性显示);(b)平行于膜平面的横截面观

■ 膜蛋白通道的转运选择性,即作用特异性,常常只是简单地依赖于企图穿过通道的分子的大小:太大了就不能通过(而有时候太小的分子也不能通过:它们与通道壁之间的间隙很窄,容不下水。参见 Finkelstein A. V., Ivankov D. N., Dykhne A. M. 2006. ArXiv:physics/0612139)。

与选择性还极为相关的是,极性基团“主动”钻进膜中非常困难,更不用说带电基团了。

你们记得,电荷 q 在介电常数为 ε 的介质中的自由能等于 $+q^2/2\varepsilon r$,其中 r 为电荷的范德华半径。很容易估算出来,对 q 等于电子电量、$r = 1.5$Å(单价带电离子的特征值),在 $\varepsilon = 80$(即在水中)时 $+q^2/2\varepsilon r$ 约为 1.5kcal/mol,在 $\varepsilon = 3$(即在仅由脂质组装而成的“纯”膜中)时 $+q^2/2\varepsilon r$ 约为 37kcal/mol。总计:自由能增量为 $\Delta F = +35$kcal/mol。根据玻尔兹曼公式,获得这么多自由能的概率,为 $(-\Delta F/kT) = \exp(-35/0.6) = 10^{-25}$。这意味着,离子向膜冲击 10^{25} 次才能成功一次。而由于“冲击”持续时间不少于 10^{-13}s(这是热振动的时间,正如我们记得的),则为了让离子钻入纯脂膜中至少需要约 10^{12}s,或约一万年……这说明纯脂膜对离子实际上是完全不可通透的。

如果在膜中埋设蛋白质,且其中形成有较宽的、充满水的通道,则离子可以沿着

通道运动，不需离开高介电系数的环境，那么就完全是另一回事了。确实，在通道中离子也将会感觉到环绕通道的膜的低介电系数，但影响会小得多：离子的自由能升高约为$+q^2/(\sqrt{\varepsilon_{膜}\ \varepsilon_{水}}R)$，其中 R 为通道半径，$\varepsilon_{膜}=3$，$\varepsilon_{水}=80$（参见 Finkelstein A. V.，Ivankov D. N.，Dykhne A. M. 2006. ArXiv：physics/0612139）。很容易估算出来，膜中半径 $R\approx 1.5$Å 的通道中离子通过时间约为几分之一秒，而 $R\approx 3$Å 时为几分之一毫秒。

离子穿过通道的通透性受到在通道中存在的位点的调控，这些位点能吸引特定的离子，从而降低这些离子需要克服的能垒。

例如，在蛋白质上靠近通道的任何正电荷的存在都会加速带负电离子的通过，而强烈地阻碍带正电离子的通过，而负电荷的存在相反，会加速正离子的通过而阻碍负离子的通过（图 12-5）。

这种效应（当然，以及孔洞的尺寸）是膜蛋白具有通透选择性的基础。

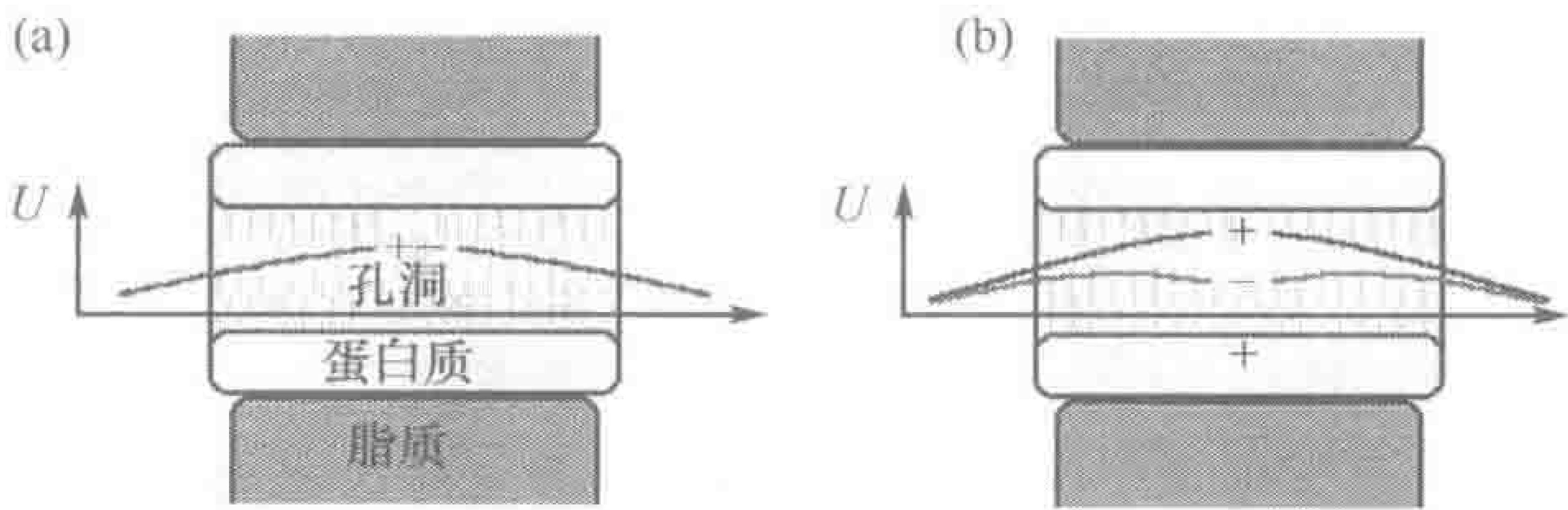


图 12-5　跨膜孔洞（图中膜垂直放置，而孔洞水平放置）和带正电离子（——+——）以及带负电离子（——−——）的静电自由能 U 的示意图。(a)孔洞内壁没有电荷；(b)孔洞内壁有正电荷

■　现在让我们着重来讨论光合反应中心（图 12-6）。它的任务是把光释放的电子转运到膜的另一侧，从而建立起光合作用工作所需的静电势。

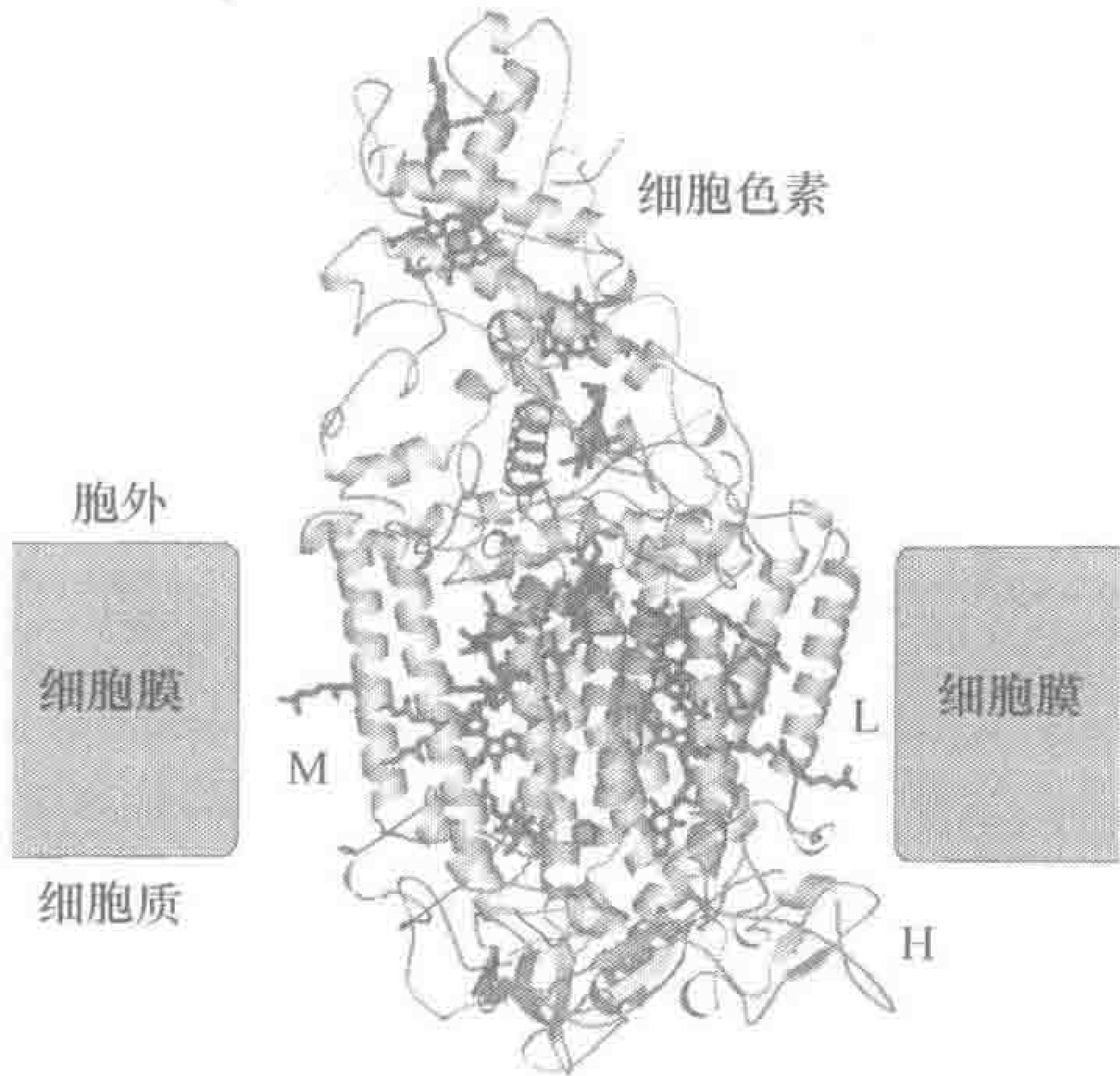


图 12-6　光合反应中心。细胞膜示意表示。蛋白质的跨膜部分由亚基 L 和 M 的螺旋，以及 H 亚基的一个跨膜螺旋组成。细胞色素完全处于外周胞质中，在膜外。请注意，蛋白质的膜内部分比膜外部分规则得多！亚基 L 和 M 结合光合色素，叶绿素和脱镁叶绿素（它们的分布在图 12-7 中看得更清楚）。这些色素和其他色素都有长长的从蛋白质插入到膜中的疏水“尾巴”。亚基 L 和 M 在它们的下面（图中）部分结合两个分子（参见图 12-7）。位于膜外的细胞色素包含 4 个亚铁血红素（和所有辅因子一样，以骨架模型显示）。图片取自文献[5]，略有修改

光合反应中心由包含 4 个亚铁血红素的细胞色素（说实在的，它不是膜蛋白：它处于膜外，在外周胞质空间中）和 3 个膜亚基 L、M 和 H（其实，H 处在膜中的仅是一个 α 螺旋）组成。亚基 L 和 M 非常相似。

所有的跨膜部分都是 α 螺旋。和所有跨膜螺旋一样，它们很长（长度等于膜的厚度），而且很规则。膜内没有不规则的环线。而处在膜外的链远没有那么有规则性，含有许多环线；大体上，它是像“普通的”水溶性蛋白（我们将在后面讨论它们）那样组装而成的。

请注意，许多嵌在光合反应中心蛋白上的相对微小的环状分子——色素：电子正是沿着由它们组成的通路“导体”移动的（电子流可以根据在电子转运时这些分子的电子光谱的变化而观察到），而蛋白质本身只是起到所谓构形绝缘体的作用。

光合反应中心又是如何工作的呢？

首先光量子从位于叶绿素上的“特殊对（special pair）”蛋白质（见图 12-7 中的示意图）上置换出一个电子。这个电子瞬间（在约 10^{-12}s 内）就跳跃到（大概是穿过相邻的“副”叶绿素 B_A）脱镁叶绿素 P_A（就是 P_A，而不会是 P_B！）上，然后，在约 10^{-10}s 内，跳跃到苯醌 Q_A 上。然后（在约 10^{-4}s 内）电子移动到苯醌 Q_B 上。为什么电子要沿这个圆形路径移动到 Q_B 上，为什么它不与“副”叶绿素 B_A 结合（根据电子光谱），仍然不清楚。

从叶绿素“特殊对”上释放出的电子被细胞色素中的亚铁血红素的电子所填补。这样就完成了反应的前半个循环。

相似的另一半循环的结果是，苯醌 Q_B 被装上第二个电子，并带二价电荷离开（带两个电荷之后，它更容易从膜上挣脱），以参与后续的光合作用。

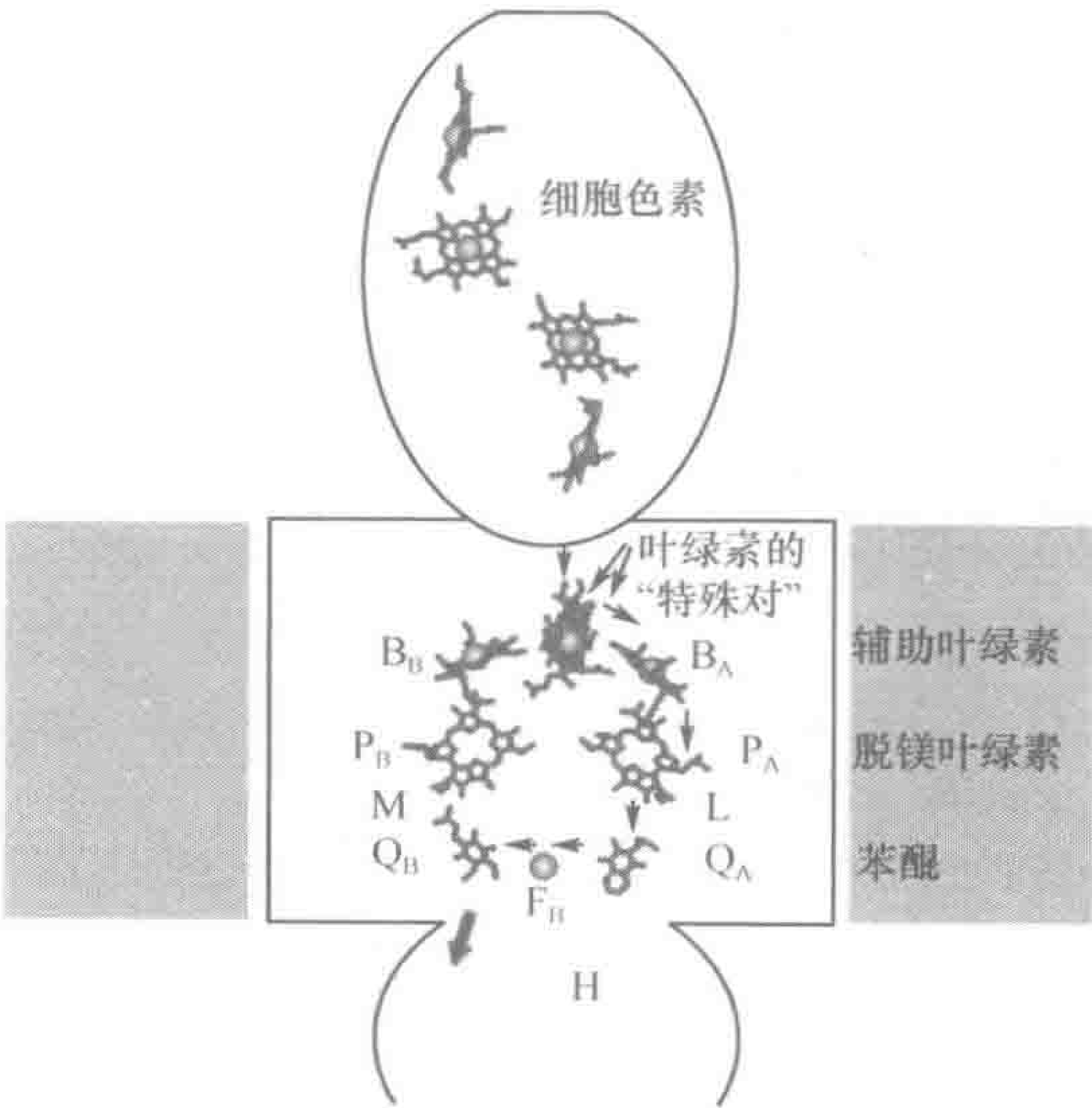


图 12-7　光合色素在光合反应中心的分布示意图。L 和 M 亚基的准对称轴穿过叶绿素“特殊对”和 Fe 离子。色素的长“尾巴”被“剃掉”，以便不塞满整个图片。电子主要是沿图上色素的与 L 亚基连接的右枝移动。电子的轨迹以短箭头表示。大箭头表示与依次到达的两个电子结合而带二价电荷的苯醌的离开。左链没有被使用。有人认为，它在过去被使用过，但现在被遗留下来了，像人体的阑尾。图片取自文献[5]，略有修改

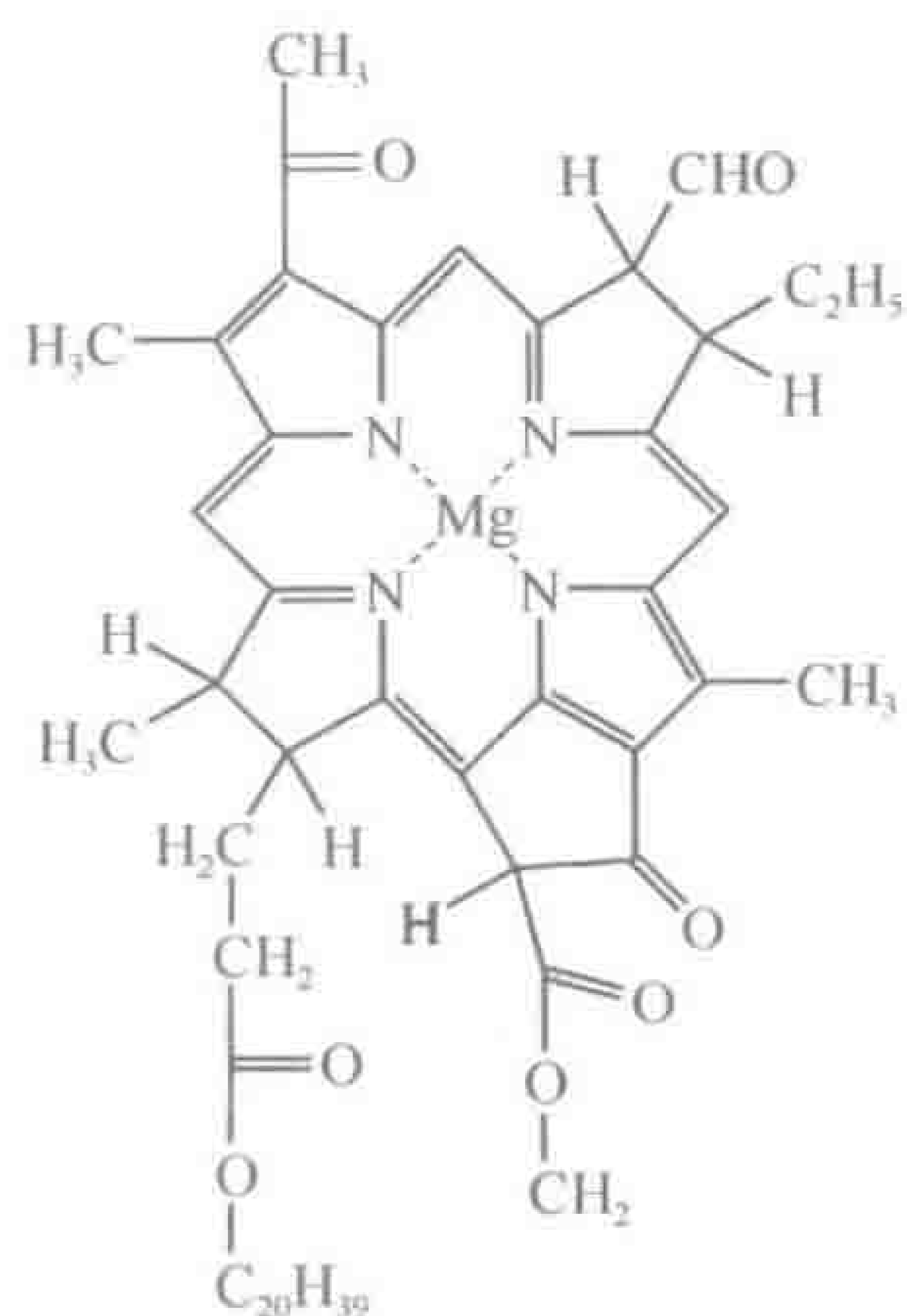


图 12-8 细菌叶绿素分子

■ 因此,光合反应中心把电子从上区室(图 12-6,图 12-7)转运到下区室,在此过程中需要反抗由转运所产生的两层之间的电势差。该反应中心的有效做功系数约为 50%(即所吸收的光的 50%被转化为被分离的电荷的能量:很不错!)。

下面强调两个重要的物理观点。

(1) 叶绿素(图 12-8)、脱镁叶绿素、苯醌及其他色素都包含有"一价半键"(…—C ═C—C ═…),即 ρ 电子键。或者说,在每一个这类分子中的电子云都是公有的(由于泡利共振:…C ═C—…⟺═C—C ═…),电子就像沿着一小块金属一样沿着这些分子奔跑。这产生了一个对电子的势阱,在该处它们"离域化",即能够移动比原子直径大得多的距离(顺便说一句,正是电子的离域化产生了色素的特征染色度:局域化于单个共价键的电子被短波长的紫外光所激发,而离域化的电子能被"普通的"波长更长的可见光所激发)。

(2) 电子从一个"金属小块"(色素)转移到另一个并不需要这些色素的直接接触。这种转移是通过量子**隧道跃迁**实现的(图 12-9)。

隧道跃迁(或者正如人们所说的,"能垒下的"跃迁:因为在这种情形下电子像是从能量能垒**下方**通过的)的本质在于,根据量子力学的定律,电子(任何粒子都一样,尤其是轻粒子)能从它所驻留的势阱的边缘稍微"溢出"。此处的"势阱"(低势能 U 的区域)就是电子所驻留的分子(叶绿素、脱镁叶绿素等)。在该电子所处的"阱"的外部(图 12-9),其势能比它在阱中的总能量(势能+动能)要高。如果没有量子效应,这种能量不足使电子不会有任何跃出势阱的可能性。但量子效应导致电子的波函数(或简单一点说是电子云密度)"溢出"势阱,虽然这个密度值随离势阱的距离增加而迅速地指数性降低。

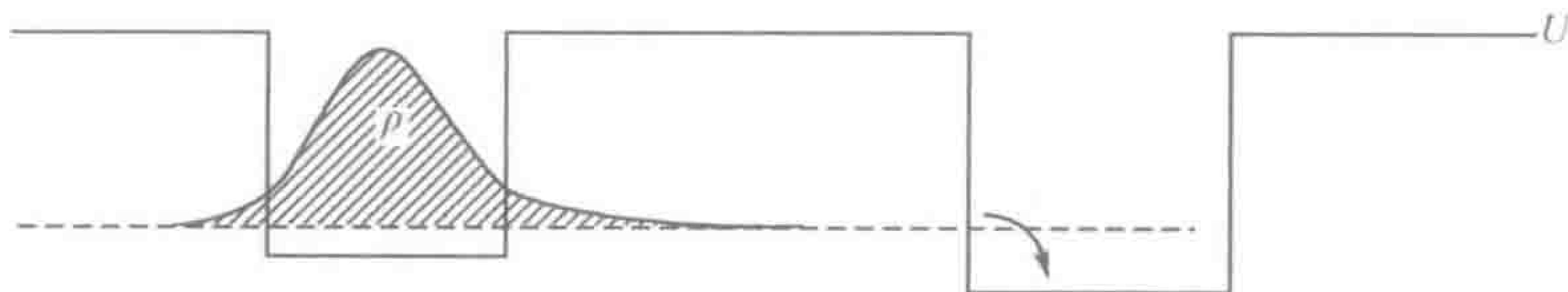


图 12-9 阐释隧道效应的示意图。粗线表示电子势能 U 曲线,虚线表示电子的总能量(势能+动能)。钟形曲线(以及下方的阴影)为电子云密度 ρ。电子最初驻留左边势阱(这是它的状态,如图所示),但它的电子云的边缘(尽管密度极小)达到右边势阱,因此电子可以随时间而全部流到那边,如果那边的能量较低的话。如果势阱深度有几个电子伏特(分子电离所需的特征能量为~100kcal/mol),则它的电子云密度降低一个数量级的特征距离为约 1Å

这也是阻止电子跌落到原子核上的量子效应:虽然这种跌落会降低电子的势能,它的动能在此情况下却会升高更多。重点在于,在电子与核间距离 Δx 趋近于

零时，电子与核的静电相互作用势能以 $1/\Delta x$ 趋向于负无穷大，而它的动能，根据海森堡**不确定性原理**而更快地（$1/(\Delta x)^2$）趋向于（在 $\Delta x \to 0$ 时）正无穷大。

实际上，按照海森堡不确定性原理，速度的不确定量 Δv 与粒子的坐标不确定量 Δx 以公式 $m\Delta v\Delta x \sim \hbar$ 相关联，其中 $\hbar$ 为普朗克常量，m 为粒子质量。换句话说，粒子在宽为 Δx 的势阱中的速度 v 的绝对值约为 $\hbar/(m\Delta x)$，粒子在某一时刻的运动方向具有完全的不确定性。因此，粒子的动能，$E=mv^2/2$ 的值约为 $m\left(\frac{\hbar}{m\Delta x}\right)^2 = \frac{\hbar}{m}\frac{1}{\Delta x^2}$。

电子处在势阱中的情况也是一样的：如果它不超出所处的势阱，则其总能量将会更高。

因此电子云很容易"溢出"势阱。在此情形下其密度指数性降低，原子的电子云密度也是如此。云密度降低一个数量级（减小到 1/10）的特征距离约为 1Å（这正是我们所知道的：原子特征半径）。

说精确一点。量子力学的计算表明电子云降低到 1/10 的特征距离 λ，为 $\ln 10 \cdot \hbar/\sqrt{8m \cdot \Delta E}$［其中 m 为粒子（即电子）质量，而能垒能量与粒子在势阱中的能量水平之差 $\Delta E>0$］，即在典型值 $\Delta E \approx 5$ 电子伏特（$\approx$120kcal/mol）时，有 $\lambda = 1$Å。

这意味着，在距离"自身势阱"3Å 处，电子云密度降低到千分之一。这说明，在一次振动中电子离开"自身势阱"到 3Å 处的概率约为 10^{-3}（距离 5Å 时约为 10^{-5}，距离 10Å 时约为 10^{-10}，依此类推）。在"阱"（色素分子）中的电子每秒进行约 10^{15} 次振动（频率在可见光范围：这很容易根据分子的吸收谱看出来）。意味着，它跃迁到另一个相距 3Å 处的"阱"（另一个色素分子）中的特征时间约为：10^{15}s/10^{-3} = 10^{-12}s；跃迁到相距 5Å 处的阱中约 10^{-10}s；相距 10Å 处的阱中约 10^{-5}s；相距 15Å 处的阱中约 1s。跃迁速度与距离的这种相关关系与在光合反应中心中观察到的结果相符。

■ 请注意一些本质性的因素。

（1）整个跃迁的距离约为 40Å。这个距离不可能由一次隧道跃迁所跨越（这样的跃迁需要的时间为 $\sim 10^{-15}$s/$10^{-40} \sim 10^{25}$s $\sim 10^{17}$ 年，整个宇宙的存在时间也不够）。然而蛋白质的构造把这个大跃迁分解为 4 个小的跃迁（从一个吸引电子的色素到另一个），而电子在几分之一微秒内跨越过这 40Å 距离。

（2）为了让电子不会立即从第二个色素返回到第一个，而是走得更远，跳到第三个以及更后面的，它的总能量（势能+动能）应该随过程的通路而降低，或者，换句话说，电子应该在每一步中从高能原子轨道跃迁到低能轨道。光合反应中心的构造要求能够满足这种在过程通路中电子在色素上的能量降低。

（3）电子在进行隧道跃迁时，并不为克服能垒而消耗能量（这里没有任何"摩

擦”)。能量的降低是由电子构象相互作用而实现的。正是因为电子跃迁到下一个色素上时,它碰到的构象恰好对应于**缺少**这个新电子的能量最小值。“新”电子到达后的能量最小值符合另一个色素的构象,它与前一个构象(即原子核分布)相比略有变形。在变到这个新构象的过程中,色素原子与环境相摩擦,多余的能量被耗散。于是电子在每一步都是从高能原子轨道进入到低能轨道,它的能量降低,能量被消耗用于“有效地”(即不可逆地)进行隧道跃迁。

(4) 隧道跃迁(或者,正如人们所说的“能垒下的”跃迁)与通常的跨能垒跃迁的活化机制的差别在于:隧道跃迁率本质上并不受温度影响(所以这种跃迁在低温下并不会消失),而活化跃迁率[正比于 $\exp(-\Delta E^{\#}/k_B T)$,其中 $\Delta E^{\#}$ 为活化能垒的能量,T 为温度]随着温度的降低而急剧减小。

第 13 讲

球蛋白。蛋白质球结构的简化表示，结构分类。β 型蛋白质的结构：β 折叠片，它们的纵向和垂直正交拼装。β 蛋白质中反平行结构占优势。折叠片右手扭转。β 蛋白质的拓扑结构。

现在我们关注的焦点集中到球蛋白，确切地说，是水溶性球蛋白。它们是研究得最深入的一类：已研究了其自发自组织的有上百个此类蛋白质，有几千个此类蛋白的原子三维结构已经被解析。因此通常在说“典型的蛋白质结构”、“在蛋白质的结构和自组织中观察到的规律”等时，指的就是它们。讲过这个必不可少的预先声明之后，让我们来看球蛋白的结构。

■ 经过在半个多世纪的广泛深入的工作，在几十个实验室中使用 X 射线结构研究（以及稍后的二维和多维核磁共振）的辛勤劳动下，已经弄清了大约 3000 种蛋白质（如果考虑到突变体和不同的功能状态，则要乘以 5 倍）的原子结构。

课堂里的声音：X 射线测定（在晶体中！）的结构，与蛋白质在溶液中的结构是相符的吗？

讲课人：一般来说，实际上是相符的（参见彩立体图对 D-17 和 D-18）。三组数据支持这一结论。首先，经常可以表明，蛋白质在晶体中保持着自己的活性。其次，有时可以在不同的晶体形式中获得相同的蛋白质，而且它的结构实际上没有变化。最后，溶液中的蛋白质核磁共振（NMR）（确实，它现在只能应用于并不很大的蛋白质）得到了与 X 射线在晶体中相同的结构。不过，应该附带说明：有时在蛋白质中具有柔性片段（某些蛋白质基团、环线、甚至在大蛋白质中的域之间的连线），因此在晶体中或被辐射源照射时它们的结构会发生变化，但这要么是蛋白质结构的微小细节。要么是域的移位，但不涉及实际上坚硬的域和小蛋白的结构。

应该补充一句，X 射线观测到不仅是蛋白质的“静态的”平均结构（这是本讲和下一讲的主题），同时还有蛋白质原子的热涨落（这将在后面简短地讨论到）。

那么，在蛋白质（水溶性球状蛋白）结构的鸟瞰图中能看到什么呢？

我们看到，较小的（由 50～150 个残基组成，少数的为 200～250）肽链折叠成直径 25～40Å 的紧密球体（图 13-1）；而体积更大的蛋白质由好几个这样的球体组成，这些球体称为“域（domain）”（图 13-2）。蛋白质链被紧凑地拼装成球体，就像有机分子堆积在晶体中一样。这可以从蛋白质表面看出来［图 13-1（a）］，也可以从蛋白质球体的“剖面”看出来，如图 13-1（b）所示。然而在讨论蛋白质时，不能光看原

子的紧密堆积的电子云(或范德华表面),否则蛋白质内部什么都看不到,而应该观察移除了原子“果肉”[图 13-1(c)]甚至是完全移除了侧基团的蛋白质分子骨架。因此不要持有那种经常由图片所产生的“蛋白质球体很疏松”的感觉。

绝大部分压紧的球体(域)的空间结构的骨架是由我们已经研究过的规则二级结构(由规则主链中的规则氢键所稳固的 α 螺旋和 β 片)组成的[图 13-1(e)]。在球蛋白中,全部 α 和 β 结构中囊括了链中 50%~70% 的残基。顺便说一句,在蛋白质分子的原子结构被解析之前,Pauling、Corey 和 Branson 已从理论上预言了这些二级结构。由这些结构组成的“堆码(stack)”[图 13-1(f)]决定了蛋白质结构的基本性质。

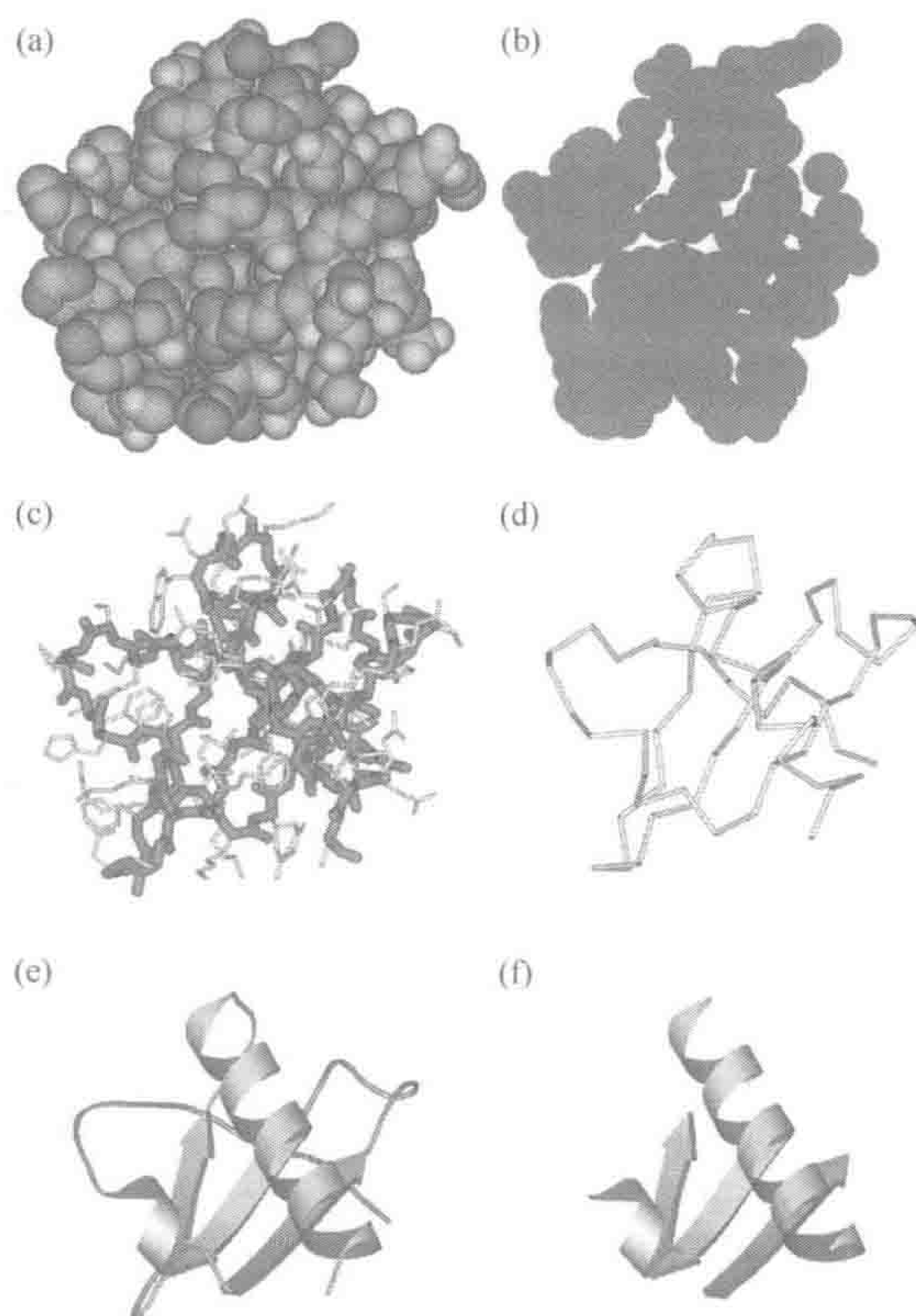


图 13-1 以不同的简化程度显示一个小蛋白[白细胞介素(interleukin)-8 的 α 亚基]的图片(参见彩色立体图对 D-12~D-16)。(a)原子模型(仅显示“重”原子,未显示氢原子);由于球体中的链紧凑拼装,我们仅能看到蛋白质的表面;(b)原子模型的剖面凸显了拼装的紧凑性;(c)主链(黑线)和侧链(较亮的突起)的骨架模型;(d)主链通路;(e)凸显主链二级结构(两个 α 螺旋和由三个 β 串组成的 β 片)的蛋白质结构示意图;(f)由二级结构组成的蛋白质结构骨架。所有图片以相同的视角和比例给出

由 α 和 β 结构组成的骨架包围着蛋白质的疏水核(或几个疏水核),而不规则

的环线则处于球体外周较远处。环线实际上从来都不会进入蛋白质球体中，这并不奇怪：那些未参与二级结构的肽基团不应该断裂自己与水的氢键，否则会导致球体的不稳定性。顺便说一句：环线（以及 α 螺旋的端部和 β 片的边缘）与水分子的氢键常常可以通过 X 射线在蛋白质分子中观察到。

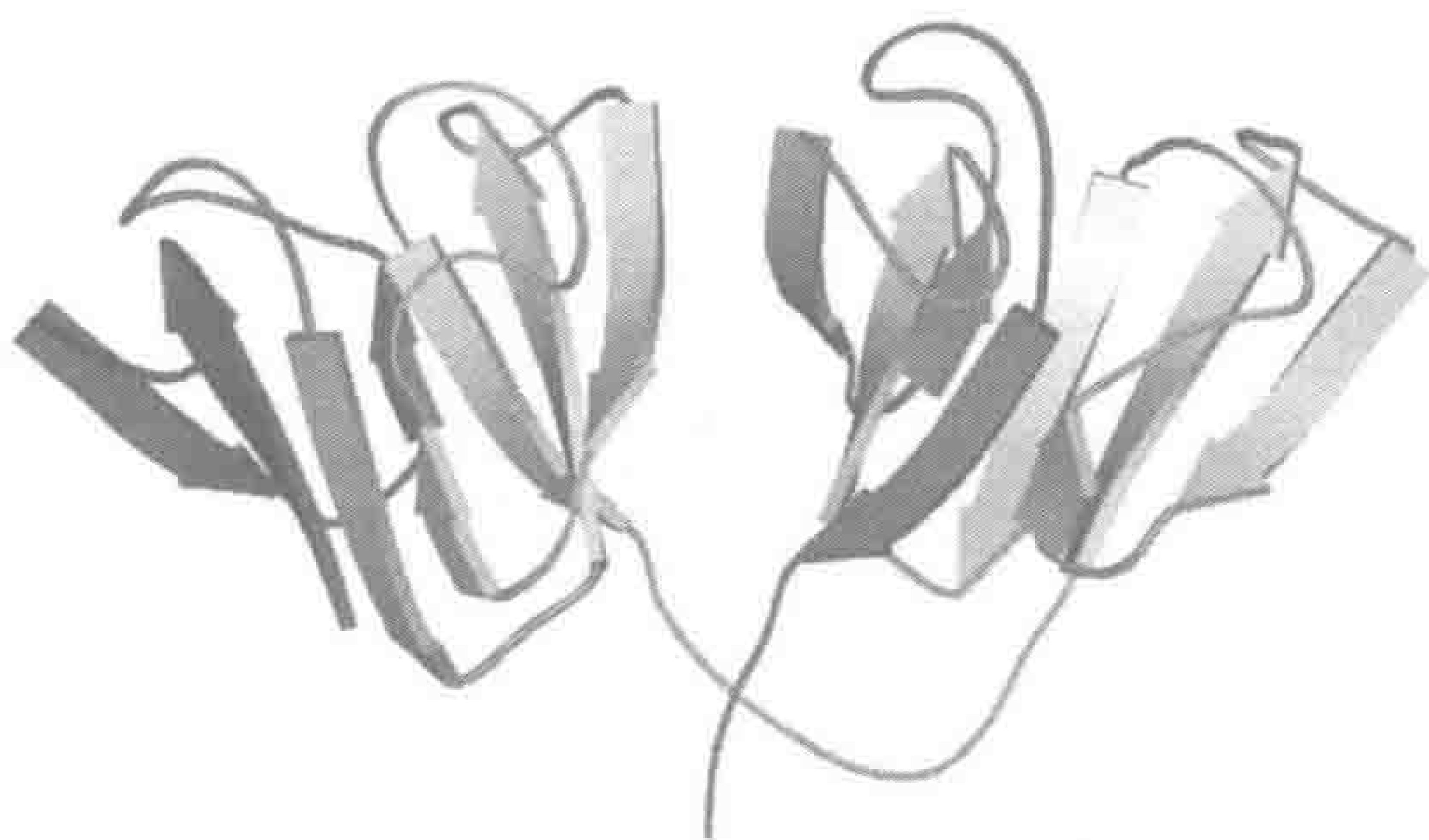


图 13-2　在 γ 晶体蛋白中的球状域

根据骨架的构造，球蛋白可以分类为："纯"β 型蛋白质；"纯"α 型蛋白质，以及"混合的"α/β 型和 α+β 型蛋白质。严格地说，蛋白质结构的这种类型分类只针对小蛋白，以及分立域（即组装成大蛋白的紧密亚球），因为大蛋白可以同时包含（例如）β 域和 α 域。

我们首先最感兴趣的是：①α 和 β 结构片段堆拼装成的紧密球体的构造[图 13-1(f)]；②肽链穿越球体的通路[图 13-1(e)]，或者如通常所说的"蛋白质球的拓扑结构"。

因此我们经常使用蛋白质结构的简化示意图（图 13-3）。这种简化不仅有助于把主要的注意力集中在二级结构（通过忽略环线的细节结构），而且忽略了这些结构的大小差别和它们的相对方位的细节。通过这种方式我们从蛋白质链的"折叠(folds)"变换到"折叠模式(folding pattern)"[图 13-3(b)]。

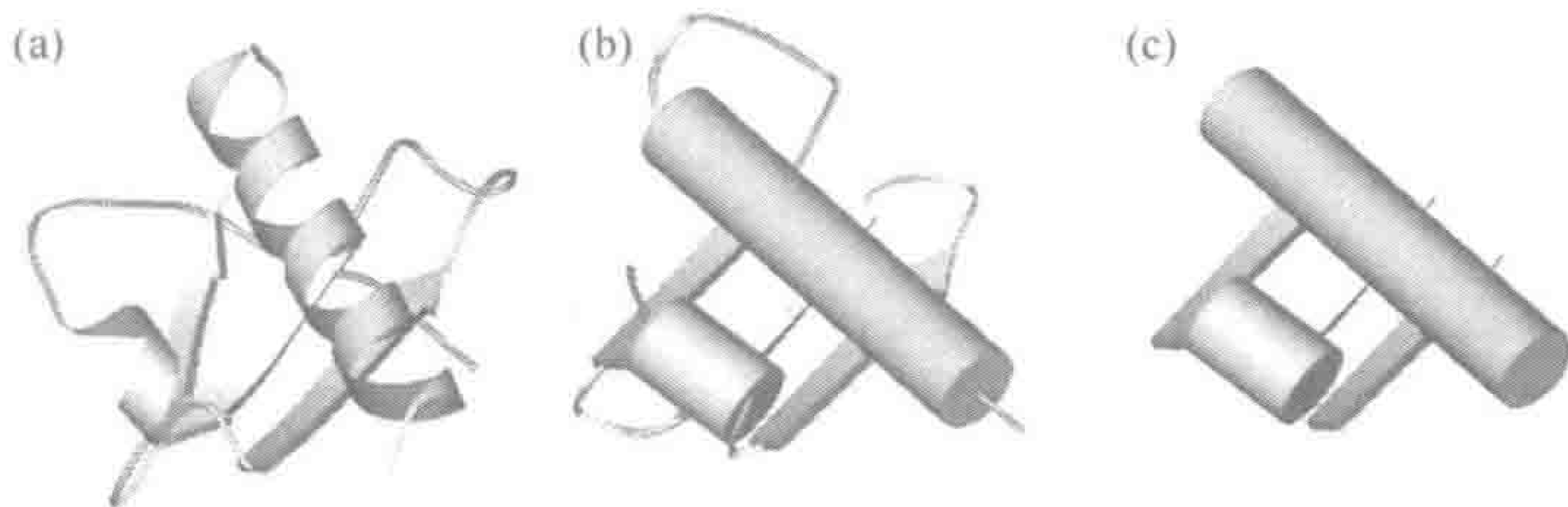


图 13-3　蛋白质结构的简化示意图。(a)描述二级结构在蛋白质链中和空间排列的**折叠**细节图[图 13-1(e)]；(b)蛋白质链的**折叠模式**：省略了环线通路的细节、α 螺旋（显示为平行的圆筒）和 β 串（显示为箭头）的大小和准确朝向；(c)结构片段的堆码(stack)：环线被移除，α 螺旋和 β 串的大小、朝向和方向都被忽略[因此，β 串被画成条带(ribbon)而不是箭头]

当我们从所讨论的蛋白质过渡到与它相当接近的（具有相似的氨基酸序列，即具有明显的共同起源）"亲戚"时，如从血红蛋白 α 到血红蛋白 β（图 13-4）时，环线结构的细节、结构片段的准确大小和朝向（甚至几个小结构片段）并不会保存，由此可以说明这种简化的合理性。

在对蛋白质球结构进行分类而必不可少的更高一级简化时，我们有时候将完

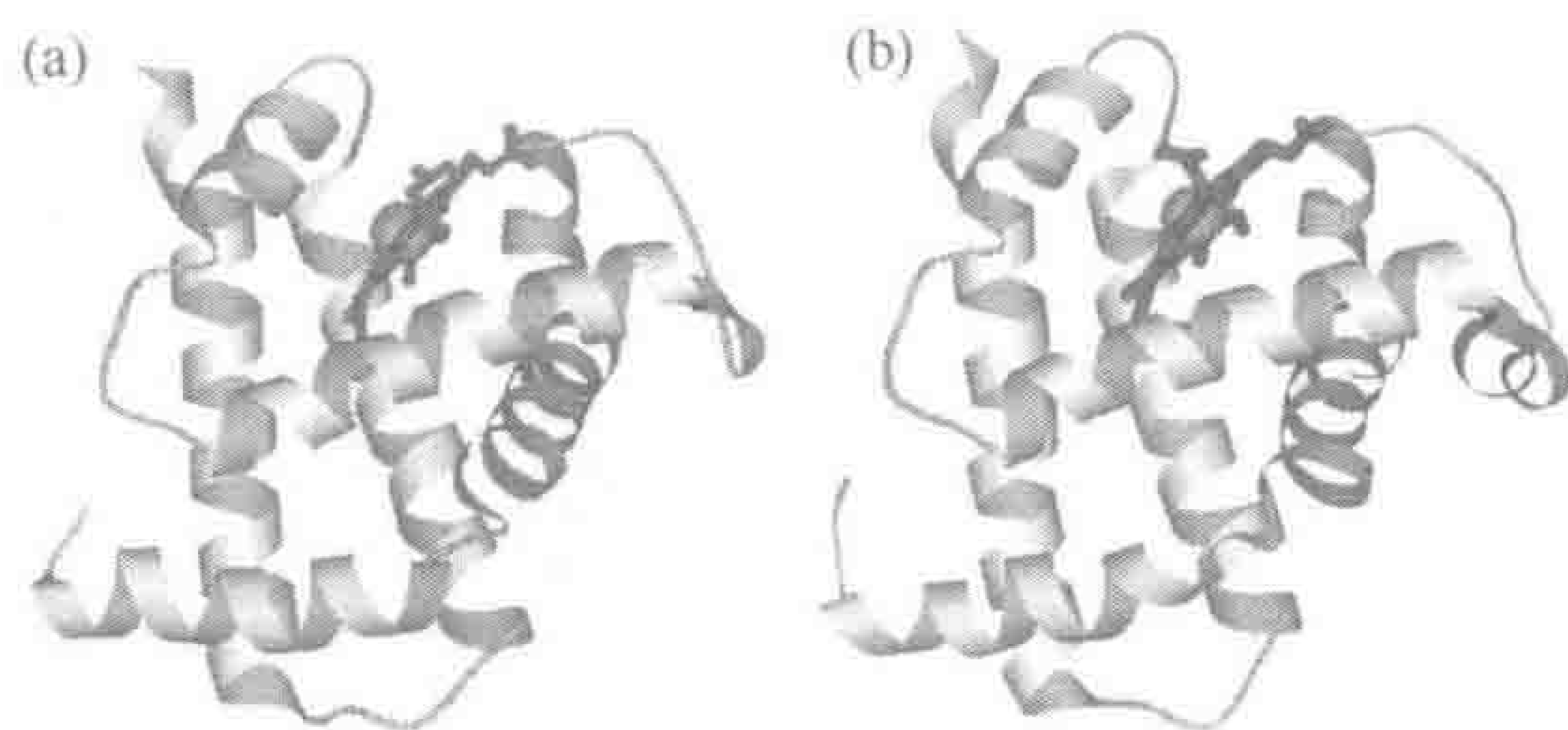


图 13-4　两个起源相近的蛋白质:马的血红蛋白 α 和血红蛋白 β(都含有血红素①)。请找出它们的相似和不同之处!(提示:总体上高度相似,差别在于环线的构象细节、几个螺旋的朝向细节,以及在右图血红蛋白 β 中多了一个螺旋圈)

全只讨论结构片段拼装成球,即由二级结构组装成的拼装包,或堆码(stack),而暂时忘记在同一分子中连接这些二级结构的环线[图 13-3(b)]。

在为了进行比较而展示由计算机给出的"真实的"蛋白质结构的同时,我将故意使用蛋白质链的折叠模式和拼装包的这些简化示意图。看上去,既然计算机能够画出"所有一切",简化示意图有什么必要呢?问题在于,"所有一切"包含了太多的细节,而示意图从中提取出主要的、在各种不同的相似蛋白质中重复出现的东西。因此它们对于蛋白质结构的分类和提取它们的主要的典型特征是很有用的。在观看蛋白质的详细图片时,我们总是需要在头脑中分析出那些典型特征,而示意图恰恰就是让我们把大家的直觉分析进行系统化。除此之外,它们有助于建立对蛋白质的"语言描述"。在实践中,当你们想要阐明所研究的蛋白质的相似性时,这些示意图以及这些语言描述对你们是用得着的:它们移除了细节,强调主要特征。当然,这些"细节"有可能是对蛋白质功能是最重要的,我们将在以后讨论——但这仅仅是强调蛋白质的功能对于其折叠模式的相对独立性。

■　肽链变成蛋白质球体的"拼装"和"折叠模式"使我们的注意力不是简单地集中在各种各样的(疏松的、透孔的等等)由结构性片段组成的复合体,而集中在由这些结构片段紧密组装成的那些复合体。因此我们勾画出对应于蛋白质链紧凑地(但没有空间重叠)拼装成球体的构象空间区域,即非价键相互作用能量最小值足够深的邻近区域。它们不仅有助于对已知的蛋白质结构进行分类,而且还可以预测新的、在自然界中尚未找到的蛋白质结构。

在这里需要加以适当说明。当我将要讲到蛋白质结构的分类、它们的相似性等时,我不是指那些无价值的东西,如所有球蛋白彼此相像,与它们工作在人体中或海洋七鳃鳗中无关之类的。这当然是真的,而且蛋白质可以根据种系发生(phylogenesis)的家族进行分类,每个家族中蛋白质的功能和(主要的是)氨基酸序列变

①heme,或译为"亚铁原卟啉"。

化并不很大。然而常见的是,具有相似的空间结构的蛋白质,在进化上(根据所有实验)却彼此互不相关。这就是我将要强调的纯粹的结构上的(而不是遗传上的)相似性。

■　让我们从研究 β 型蛋白质的结构开始。看来,β 结构的域的构造比其他的域更简单:伸展的肽链片段形成两个(有时是几个)β 片,并互相组合起来。换句话说,在 β 型蛋白质中,二级结构的"堆码"看上去相当简单。

在 β 型蛋白质中占优势的是**反**平行 β 结构。

由于蛋白质是由不对称的 L 型氨基酸构成的,所以伸展的 β 串稍微地沿自身扭转:正如你们记得的,伸展构象的能量最小值位于拉氏图的对角线上方。片中扭转的 β 片段在氢键的作用下联结起来,结果这些 β 片也以特定的方式扭转(图 13-5)。它们的表面就像螺旋桨。β 片中相邻的伸展小段之间的角度约为−25°。这意味着,如果横着 β 串的通路看过去,这个螺旋桨看上去是**左**旋的[图 13-5(a)]。如果顺着 β 串通路看过去,则这种扭转是**右**旋的[图 13-5(b)]。通常人们都是顺着 β 串通路看过去,并且说,β 片具有**右**旋的扭转性。

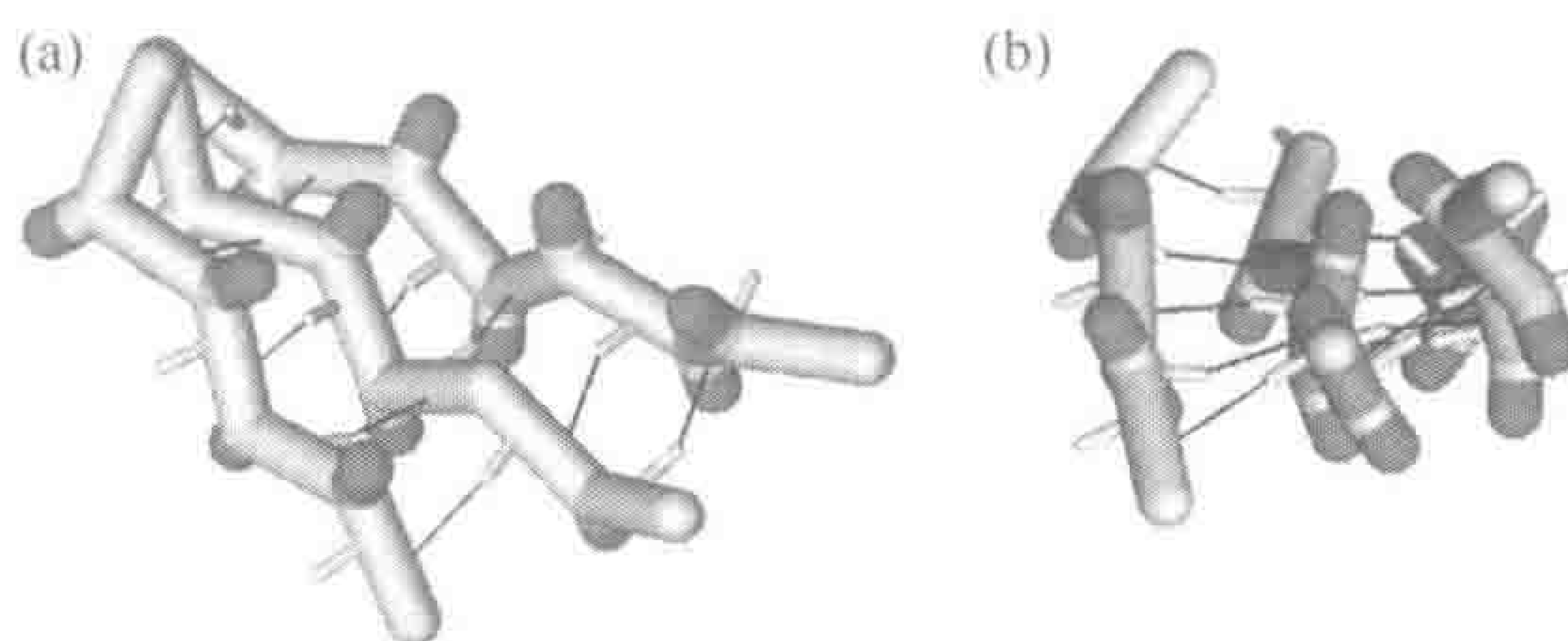


图 13-5　β 结构片,横着 β 串观(a)和顺着 β 串观(b)。β 片具有褶皱的表面(向外伸出的 C^{β} 原子凸显了这一点),而且(通常)具有右旋的(顺着 β 串看过去)扭转性。β 串之间的氢键以细线条表示

■　存在有两种基本的 β 片拼装类型:正交拼装和纵向拼装[图 13-6(a),图 13-6(b)]。在两种情形下 β 片都是"面对面"拼装,域的疏水核包含于它们中间,但在前一种情形下 β 片彼此相对扭转强烈:角度约为 90°(±10°~15°),后一种情形较弱,−30°(±10°~15°);处于这两个范围之外的角度(尤其是,约+30°的角度)很少观察到。

在正交拼装[图 13-6(a)]中,β 片段有扭转且通常有点弯曲,致使总的"堆码"形状像个圆筒,圆筒轴与 β 串之间有很大倾斜。这种拼装通常被称为 β 圆筒或 β 圆桶,虽然在由反平行 β 结构构成的 β 圆筒中(与后文要讲到的由平行 β 结构构成的 β 圆筒不同),氢键网络常常在小桶的相对侧面处断裂(或拉开),这使得区分两个 β 片相当容易。在这种拼装的"闭合"角处,两个片层的 β 片段非常靠近,使得肽链从以代价高昂的 90°转弯从一个片层中走到另一个片层中。可以说,是同一个 β 片弯曲并架在自己身上。在对面的"张开"角处,β 片张开,所形成的裂缝通常由 α 螺旋或不规

则环线所填充,或者是由活性中心所填充,如在视黄醛结合蛋白中那样(图 13-7)。

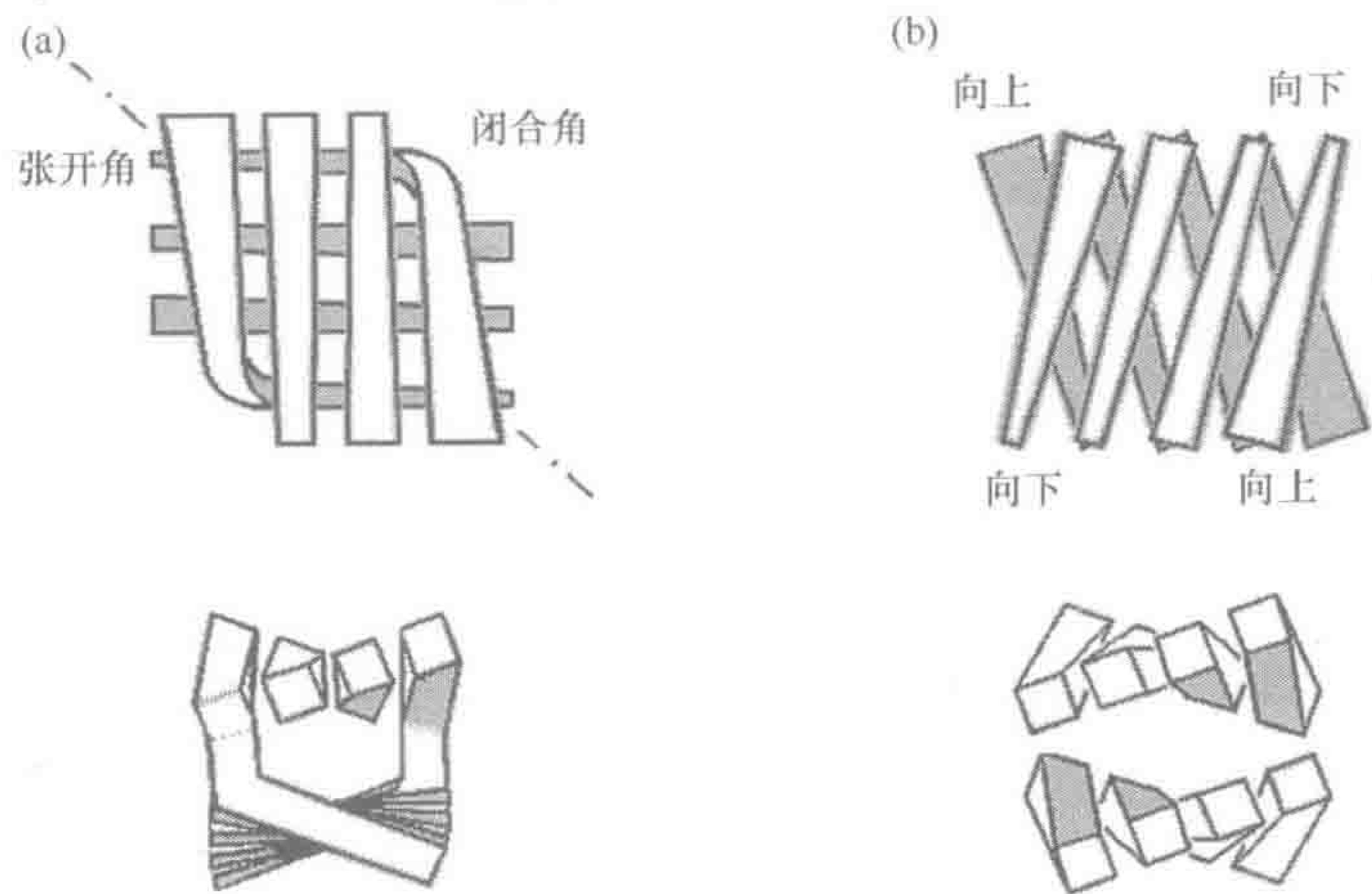


图 13-6　β 片的正交拼装(a)和纵向拼装(b)。β 片拼装的侧视图(上)和下视图(下)。在侧视图中,β 串的变宽表示它向读者接近。虚线表示正交 β 圆桶的轴。该拼装的两个"张开"角处于轴上。该处 β 片张开得最大。在另外两个"闭合"角处 β 片最接近;此处肽链弯曲并从一层转向另一层。在正交拼装中疏水核具有近似于圆柱的形式。相反,在纵向拼装中,疏水核是扁平的,扭转的片层之间的距离变化很小,一个片层相对于另一个片层的转弯使彼此平铺的扭转 β 串的疏水表面在长距离上保持接触。图片取自 Chothia C. , Finkelstein A. V. , 1990. Annu. Rev. Biochem. 59:1007-1039,略有修改



图 13-7　视黄醛结合蛋白:β 片正交拼装的例子。肽链通路好像"河曲(meander)"图案[见拓扑示意图(β 圆筒的扁平展开图)侧面观]。在该示意图中 β 串以箭头表示。之所以称为"河曲",是因为沿肽链相邻的 β 串在圆筒表面上相邻;它们是由氢键相结合的(存在于边缘的 β 串 1~8 之间的键,在扁平展开图上以小黑点表示)。视黄醛结合中心处于圆筒轴上。视黄醛以骨架模型显示。拓扑示意图上的数字表示结构片段在肽链上的顺序

■　纵向拼装[图 13-6(b)]是那些仅有螺旋桨似的扭转,而不以弧线弯折的 β 片的典型拼装形式。这种拼装通常称为 β 夹层。β 夹层的两端被不规则的环线所覆盖(可在图 13-8 中看到)。在某些 β 夹层中,β 片的边缘串非常接近(有时它们甚至以几个氢键相连接),以致该拼装获得了圆筒形,在圆筒轴和 β 串之间存在有微

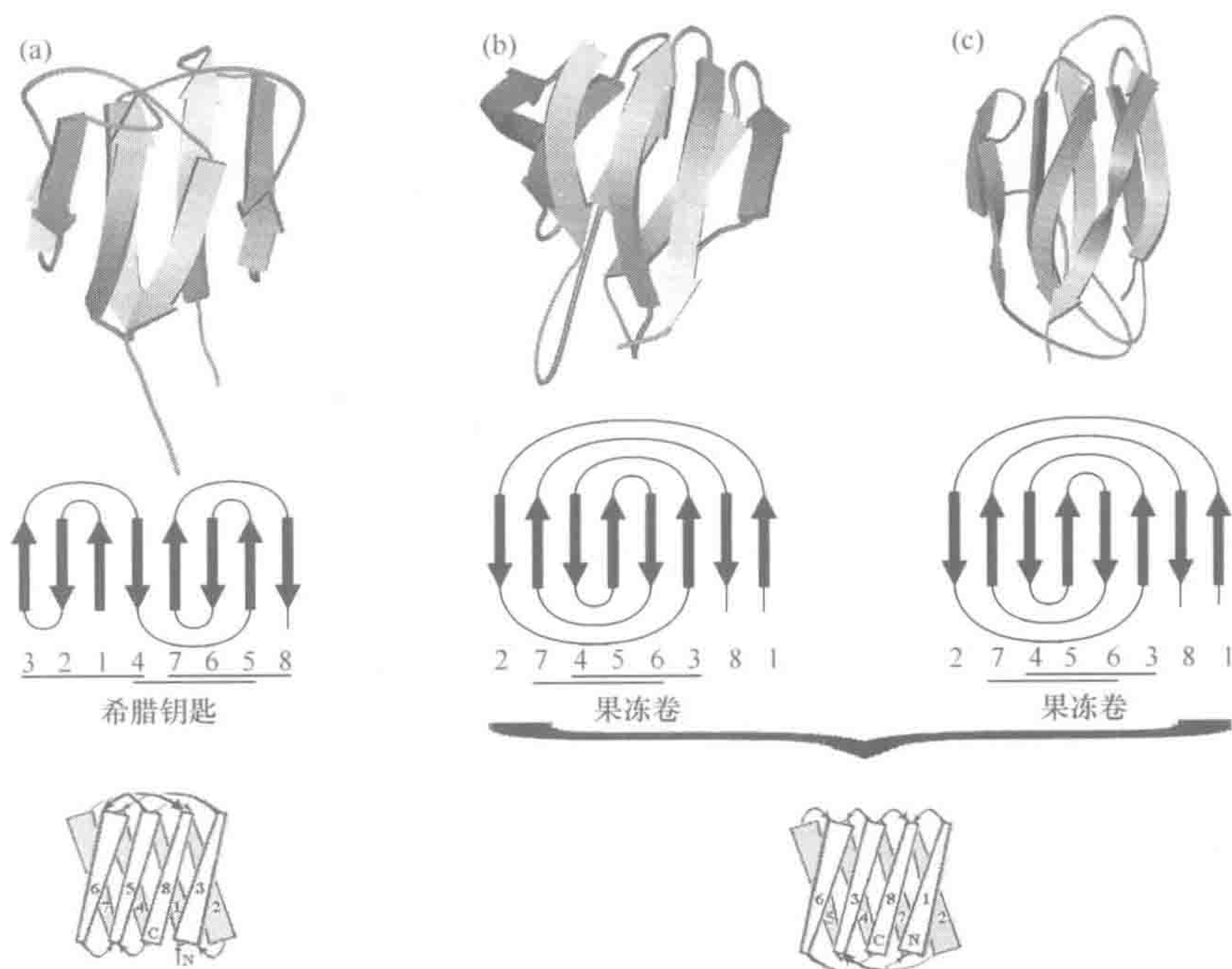


图 13-8　β 片纵向拼装的例子。在(a)γ 晶体蛋白的一个域中(图 13-2)、(b)分解代谢活化蛋白的蛋白质 β 域和(c)卫星烟草坏死病毒外壳蛋白中的肽链折叠模式。这些蛋白质的拓扑示意图显示于下方。所有这些蛋白质包含有典型的"希腊钥匙"(或"β 发夹等分弯曲")的拓扑结构,其中在肽链上相邻的四个 β 串反平行,而且第一个和第四个以氢键相结合。在拓扑示意图上两个 β 夹层展开(就像侧面切开的圆筒)并摆在平面上。切开位置选择在能突出显示肽链折叠对称性的地方。于是在 γ 晶体中切口在 β 串 3 和 8 的接头处,就突出显示了该域的前一半(β 串 1~4)和后一半(β 串 5~8)的相似性。示意图中 β 串的靠近显示了它们之间的氢键;展开图边缘处的氢键(如果有的话)以边缘处的小黑点表示。β 串之间增大的间隙使 β 片夹层相互分开(如果它们之间没有氢键)。γ 晶体蛋白的域含有由 β 串 1~4 和 5~8 构成的两个重复的希腊钥匙;以及一个由 β 串 4~7 构成的希腊钥匙。在图(b)和(c)中显示的蛋白质包含由 β 串 3~6 和 4~7 构成的"钥匙"。除此之外,它们拥有多次弯曲的 β 发夹的拓扑结构(通常称为"果冻卷"):其中不仅像"希腊钥匙"那样典型地有氢键连接 β 串 3 和 6、4 和 5,还有额外的氢键连接 β 串 1 和 8,2 和 7。请注意,"希腊钥匙"拓扑结构比"河曲"(图 13-7)能让蛋白质链更好地围绕球体的核:除了用带侧枝的 β 结构之外,还用上方和下方的环线。通常,β 片的纵向拼装是形成 β 夹层[(a),(b)],但其中的某一些(例如,卫星烟草坏死病毒的外壳蛋白和一系列其他病毒的外壳蛋白)可能形成带有共轴 β 串的 β 圆筒(c)

小的夹角。

必须强调,蛋白质链的折叠模式比已知的蛋白质球要少得多,而"堆码"又比蛋白质链的折叠模式要少得多。同样一个"堆码",即同样的片段结构的拼装,可能对应于不同的从链到球的折叠模式,即这些片段可能以同一个多肽链按不同的方式相连接。图 13-8 显示,同样一个 β 夹层(按不同的拓扑,即链穿过这个夹层的

不同途径)是如何用来作为三种不同的蛋白质域的基础结构的:(a)γ 晶体蛋白的一个域,(b)分解代谢活化蛋白的 β 域,以及(c)病毒外壳蛋白。而且后两个蛋白具有甚至同样的肽链折叠模式,即在它们中肽链完全相同地穿过相同的 β 结构堆码(在图 13-8 中以大括号强调)!

■ 在蛋白质之间缺乏任何其他可以看得出来的联系的情形下,这样的结构相似性是非常之多的。

在这里我忍不住要向你们展示另一个以 β 夹层为基础结构的蛋白质。这就是免疫球蛋白,更准确地说,这个大蛋白质的每一个域都是这样(或者大致是这样)构成的(图 13-9,参见彩色立体图对 D-22)。

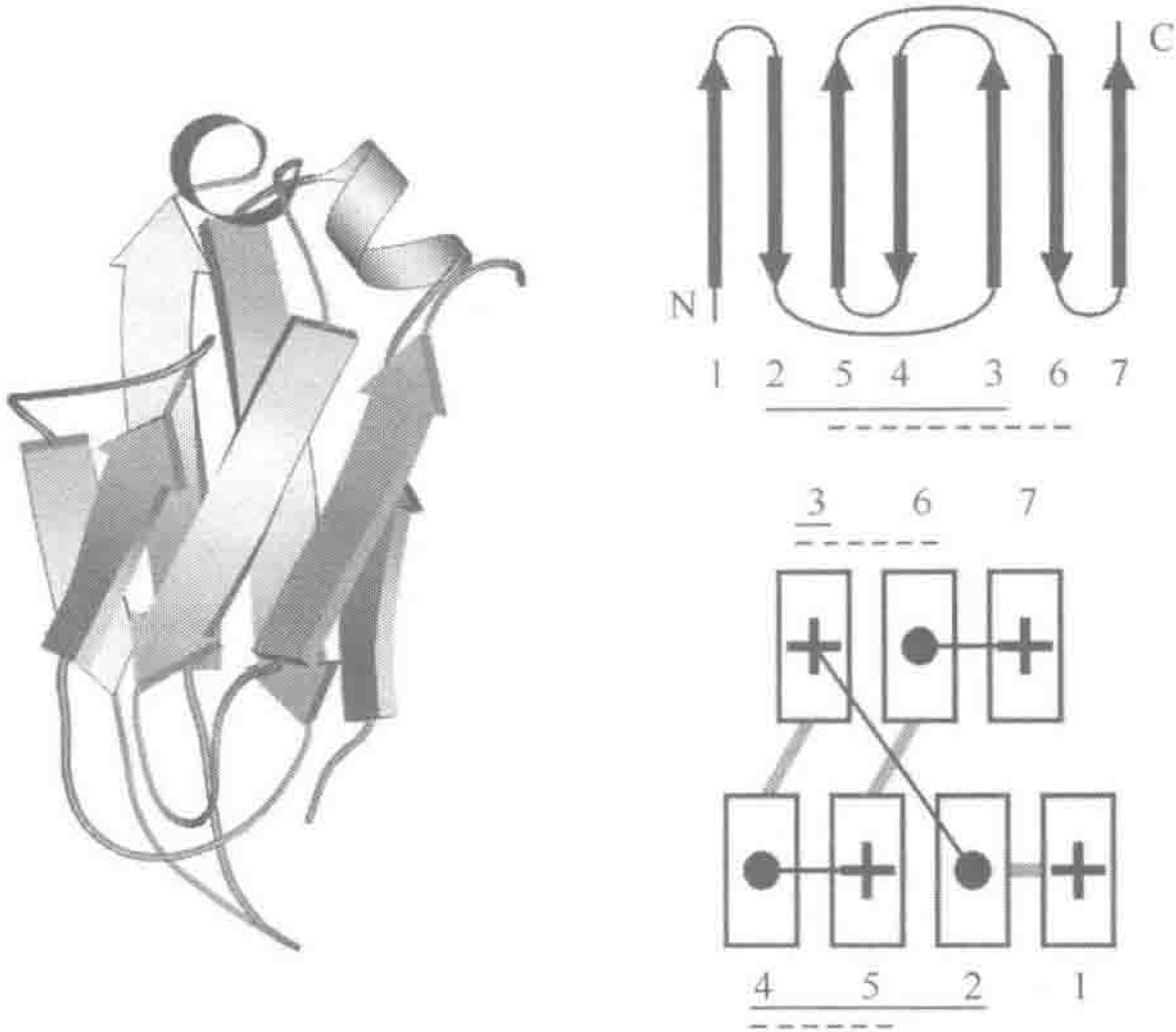


图 13-9 在免疫球蛋白 κ 的不变域中的 β 片纵向拼装。左边显示了该蛋白质的完整图,在拓扑示意图(右上)中突出显示了“希腊钥匙”。在右下图中给出了“从下方”(从结构片段的顶端)看过去的样子。矩形:β 串的顶端。十字对应于片段的 N 端(即它“离开我们”),点表示 C 端(即它“走向我们”)。连接结构片段的环线的通路显示为黑线(如果环线靠近我们)和灰线(如果环线处于折叠的对面)。注意,这样的示意图能以最简化的方式展示这些片段(β 串)的纵向拼装。除此之外,它还能让人们看见“希腊钥匙”的空间构造并觉察到,空间中所具有的两个“钥匙”(分别由串 2、3、4、5 和 3、4、5、6 组成)是按不同方式组织起来的。由两个交叠的(如本图中所示的)希腊钥匙组成的结构有时被称为“完整的希腊钥匙”

大约有其他 50 多个与免疫球蛋白在氨基酸序列上并不相似(虽然它们中的一部分也像免疫球蛋白一样负责与某种对象的特异性结合,如在细胞间的识别过程中)的超家族(它们内部之间在一级结构上**没有**相似之处,但折叠模式相似)的蛋白质链都以相似的方式组装起来。尽管在某些这样的超家族中,还观察到与图 13-9所示的“标准”结构有微小的区别。

例如,在某些蛋白质中,β 串 1 与 7 形成平行 β 结构(甚至可能失去与串 2 的

连接);在另一些中,β 串 4 从 5 移到 3;在第三类中,在 β 串 3 与 4 之间的连接带中形成了额外的 β 发夹。但是由 β 串 2、3、5、6 和 7 形成的折叠骨架维持不变。

在愉快地向你们展示了这一非常流行的肽链折叠模式之时,我还有一个目的:图示那些含有结构片段纵向拼装的蛋白质的构造模式时,最简单的是以表现这类蛋白质的顶端查看视图的示意图表示出来。

回到 β 片的正交拼装。图 13-10 显示,胰凝乳蛋白酶型或胰蛋白酶型的丝氨酸蛋白酶(a)和胃蛋白酶型酸性蛋白酶(b)的共同点是具有相同的 β 圆筒(但具有不同的拓扑,肽链穿越 β 片正交拼装的通路不同)。

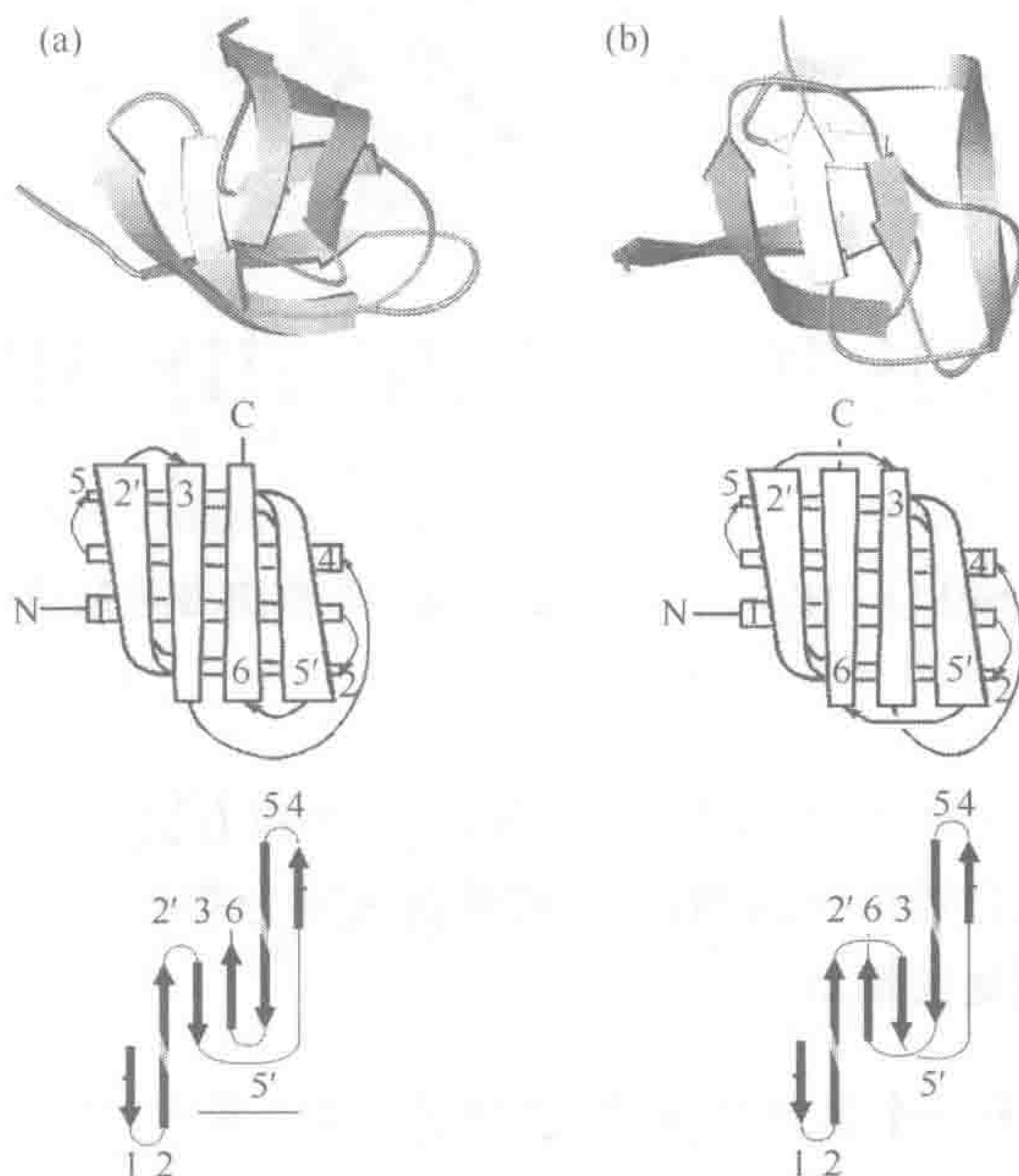


图 13-10　胰凝乳蛋白酶型丝氨酸蛋白酶(a)和胃蛋白酶型酸性蛋白酶(b)中的肽链折叠模式。在后一个蛋白质中环线被缩短并相当示意性地绘制。图中显示了这些蛋白质中的 β 片正交拼装示意图和 β 片的拓扑示意图。这些 β 片有弯曲(在弯曲处,片的边缘远离读者并紧接在一起;在此出现的 β 片边缘有氢键显以短线表示);β 片的折弯处对应于它们的较亮部分。胰凝乳蛋白酶中的"希腊钥匙"以下方的横线强调

■　上面我们讨论了 β 型蛋白质的最有意义的"基本"构造。还有其他"基本构造",例如,"多叶螺旋桨"。

在神经氨酸酶(neuraminidase)①(图 13-11)的"螺旋桨"中,6 个倾斜地组装的 β 片形成了玫瑰花形(在其他蛋白质中可能多至 8 个 β 片)。如果把这些 β 片认为是成对的,则它们形成了 β 夹层,因此"螺旋桨"经常被描绘成是由 β 夹层组装成的超圆筒。

①神经氨酸酶又称唾液酸酶,是分布于流感病毒被膜上的一种糖蛋白,它具有抗原性,可以催化唾液酸水解,协助成熟流感病毒脱离宿主细胞感染新的细胞。



图 13-11　神经氨酸酶中“6 叶螺旋桨”形的 β 结构。下图：该蛋白质的拓扑示意图，由 6 个反平行 β 片组成。图片取自文献[5]，略有修改

在这个超圆筒的轴上“凹部”（我们看到，它并没有被环线所遮盖）包含着活性中心。在视黄醇结合蛋白中，我们已经碰到过这样分布的活性中心：处于圆筒的凹部。我们还会见到这类情形。

■　“β 棱柱”或（另一个名称）“β 螺旋”型的构造（图 13-12，参见彩色立体图对 D-24 和 D-25）同样令人感兴趣，它很有规则性。这种棱柱的 3 个棱面由 3 个 β 片组成（而且是平行的！），而肽链按螺旋沿它们走过，连续地由一片走到另一片。肽链看上去就像是在棱柱轴上盘旋，或是在平行 β 串的连接下形成常见的右旋螺旋，或是（在其他棱柱中）左旋螺旋，这在 β 串连接中特别少见。

值得注意的是，在某些原始鱼（如七鳃鳗）中，免疫球蛋白的功能不是由以经典的免疫球蛋白折叠方式构成的蛋白质行使的，而是由以 β 棱柱构成的蛋白质行使的。

■　现在是合适的时候来讨论 β 型蛋白质的拓扑结构。首先注意，β 型蛋白质主要是由反平行 β 结构组成的。我们至今为止所讨论的 β 型蛋白质大部分是由纯粹的反平行 β 结构组成的。有时其中掺杂有少量的平行 β 结构[图 13-10(b)]。由纯粹的平行 β 结构组成的蛋白质也存在有，但是非常罕见。

平行 β 结构和反平行 β 结构的混合型是相当少的，这并不奇怪，因为它们具有几种不同的构象，以至于彼此的连接处应该是能量不利的。至于结构的能量不利性与其出现频度的相对罕见性在何种程度上相关，我们将在最近的一讲中去讨论；

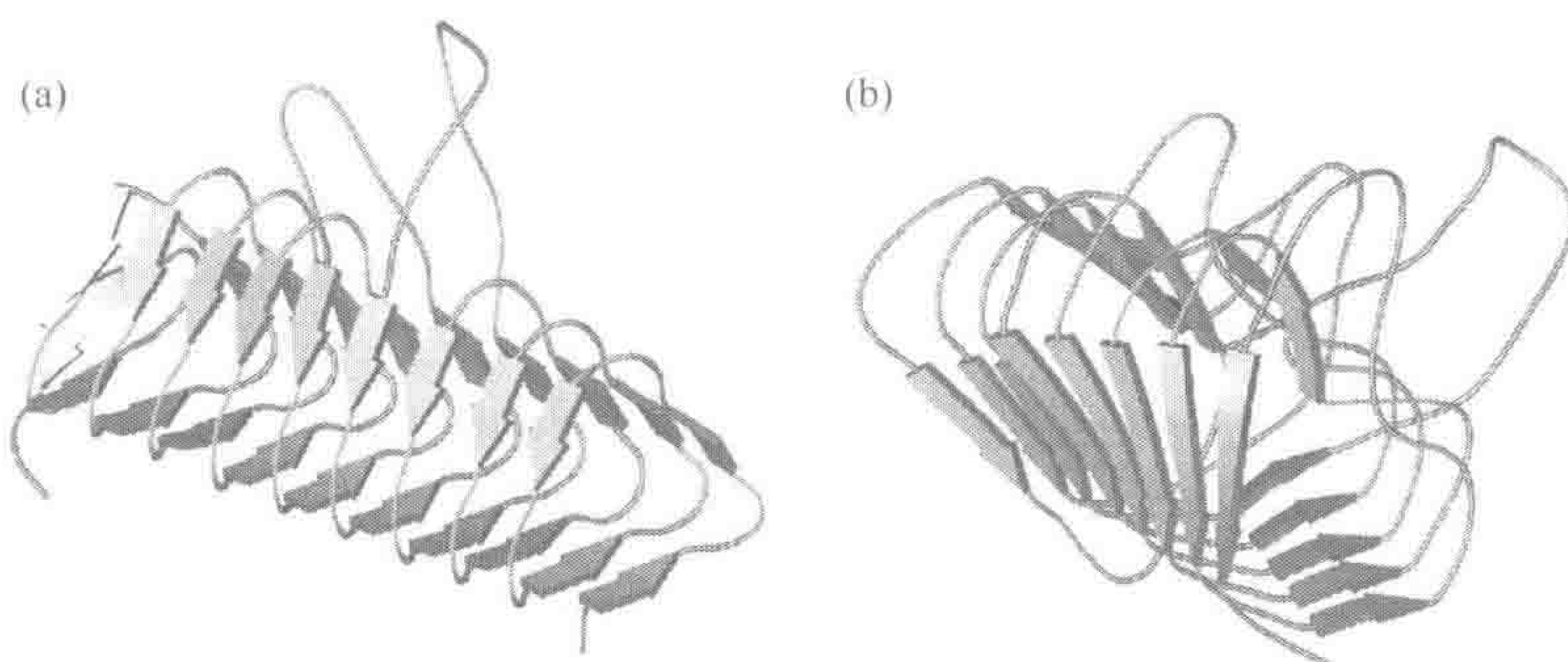


图 13-12 酰基转移酶(a)和果胶裂解酶(b)中的β棱柱。注意肽链绕棱镜柱轴的不同扭转性:(a)中特殊的左旋和(b)中常见的右旋,还请注意在右旋方式下肽链不存在常见的β片扭转性[沿β串看过去的右旋螺旋桨扭转。图 13-12(b),图 13-5~图 13-11]

但是总体上很清楚,稳定的系统(蛋白质是稳定的)应该基本上是由同样稳定的部件所组成,而且应该避免内部的不稳定部件。

β型蛋白质中β片的**反**平行特征占优势,这与它们的构造通常基于β发夹(图 13-13)而密切相连。这种发夹经常对半弯折,有时候两次弯折甚至三次弯折,如图 13-8(b)和图 13-8(c)所示。



图 13-13 反平行β发夹。弯折的发夹对β型蛋白质的边缘是很典型的,“倒钩对”对这些蛋白质的中间很典型的

连接β片段的环线,进口和出口通常在同一折叠的边缘(即它们并不贯穿“堆码”,而是遮盖住其顶端)。这几乎可以在所有图中都看到。而且环线,甚至很长的环线,也总是连接在空间上彼此接近的β片段的末端。因此沿肽链相连的β片段通常**不是**平行的,而是常常形成反平行的β发夹。

还要指出的是,环线在环线上的“交叠”(或“环线交叉”)极少碰到[这一原则的一个例外显示于图 13-10(b)下方],想必是因为在这种情形下环线需要更加弯曲(以避免碰撞或脱水),这也是能量不利的。避免环线交叠是在蛋白质中观察到的一般性结构原则。

结果就是,在图 13-14 所显示的由同一个肽链段组装成的大量可能的β片构象中,在目前经常碰到的只有三种模式:“河曲”和两种“希腊钥匙”(图中以下划线标出),正是这些模式没有上面所提到的所有不利因素。

顺便说一句,“meander”(中文意译为“河曲”)是在希腊的一条非常蜿蜒曲折

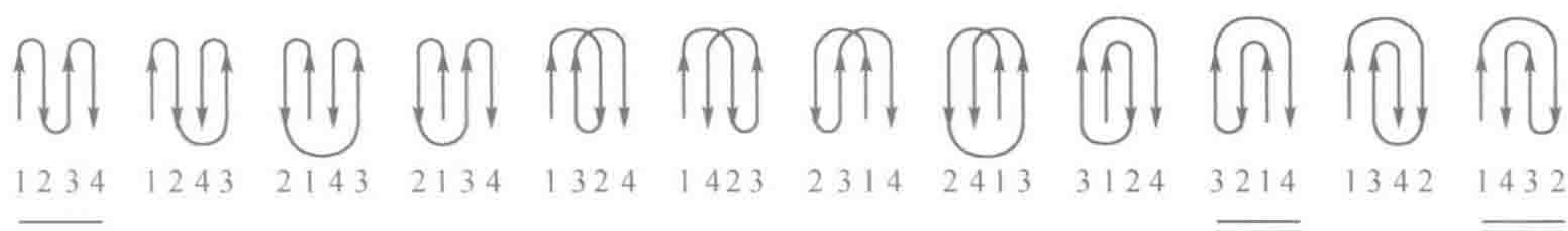


图 13-14　由 4 个 β 片段构成的 β 片的可能拓扑结构。图中只显示了那些两个连续(在肽链上)走向的 β 片段指向相反方向的 β 片。在这些拓扑结构中常见的只有“河曲”(下方单下划线)和两种(双下划线)“希腊钥匙”(其区别在于,串 1 是处于“钥匙”内部或在它的边缘)。有河曲的蛋白质的例子:视紫红质结合蛋白(图 13-7);含有“希腊钥匙”的蛋白质的例子显示于图 13-8~图 13-10(而且同一蛋白质中常常含有这两种类型的“钥匙”)。注意,所有带有“混合”(反平行和平行)β 结构和带有环线交叉的拓扑结构都极少碰到

的河流的名称。其特征在于,沿肽链相邻的 β 串同时也是在空间上最近的邻居(图 13-8);通常(但并非永远)它们以氢键相连。“希腊钥匙”(这种图案装饰可以在古代花瓶和花园栅栏上看到)的特征在于,四个沿肽链相邻的 β 串呈反平行,而且第一个和第四个以氢键相连。而且实际上“钥匙”的第二个和/或第三个 β 串常常并不是处在同一个 β 片上,如图 13-14 中所示,而是在另一个 β 片上。因此它们以相同的希腊钥匙拓扑结构可形成各种不同的空间结构,即所谓的 Efimov **“abcd”结构**[①](图 13-15)。请在图 13-8 和和 13-10 中找出它们来。

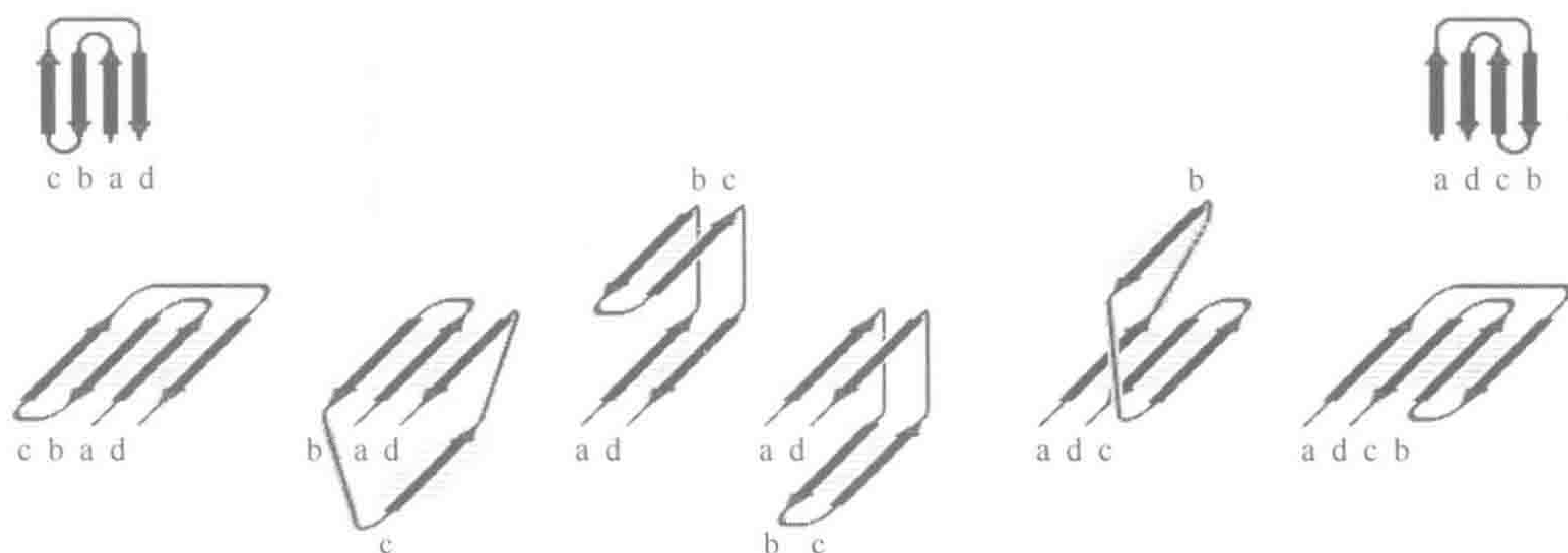


图 13-15　具有相同的希腊钥匙拓扑的各种不同的 Efimov“abcd”空间结构。注意由同一个 β 片的两个平行 β 串和处于肽链上两者之间的另一个 β 片的 β 串所组成的超螺旋(即下排左二由 β 串组成的超螺旋 b-c-d 和下排右二的超螺旋 a-b-c)的右手螺旋通路[②]。这种在同一个 β 片中的平行 β 串的“右旋”连接对蛋白质是很常见的;相反,“左旋”的连接很罕见。

■　所有这些典型的、在蛋白质结构中经常碰到的结构(发夹、河曲、希腊钥匙、abcd 结构等),都是由沿肽链相邻的 β(和/或 α)结构组成的,常被称为“**超二级**结构”。

---

①俄罗斯科学院蛋白质研究所研究员 Alexander V. Efimov 以 abcd 结构为依据构建了蛋白质超家族的结构树。参见网站:http://strees. protres. ru/.

②右手半握,拇指向前伸出。四指指向 β 串转动方向,则拇指方向与第一个和第三个 β 串的方向相同。

第 14 讲

α 型蛋白质的构造。螺旋束和螺旋层。α 螺旋构成的准球模型。α 螺旋接触下的紧凑拼装。α/β 型蛋白质构造:α 螺旋覆盖的平行 β 折叠片,以及 α/β 圆筒。β-α-β 亚基的拓扑结构。α+β 型蛋白质的构造。蛋白质的架构与其功能之间并无直接联系。

现在来讨论 α 型蛋白质:由 α 螺旋构成的蛋白质。它们比 β 型蛋白质更难分类。问题在于,在 β 型蛋白质的片层中,β 串的相互位置排布是由主链(在各处都相同)中的氢键所维系的,而 α 螺旋在球体中的位置排布是由其侧链基团的紧凑拼装所维系的,但侧链基团在尺寸上类型多样、差别极大。因此 α 螺旋不能像 β 串那样拼装成比较标准的片层。

■ 由长 α 螺旋构成的 α 型蛋白质的构造要比其他的 α 型蛋白质简单一些。这些螺旋彼此(大致上)平行或反平行地(简而言之,“共轴地”)排列,构成螺旋束。我们已经在纤维蛋白和膜蛋白中见过螺旋束了。

在图 14-1 中展示了 3 种四螺旋束的 α 型蛋白质。这些彼此相似的蛋白质行使不同的功能:第一个是细胞色素,结合质子;第二个(蚯蚓血红蛋白)结合氧;而第三个(烟草花叶病毒的外壳蛋白)所结合的分子大得多:其他外壳蛋白和 RNA。前两种蛋白质的功能可能具有某种共同之处,因为它们都是在呼吸链中运载某种东西。不过,这里的相似性非常远:在细胞色素中,多肽(我们讨论的正是它的折叠)结合血红素,而血红素结合携带电子的铁离子;在蚯蚓血红蛋白中,多肽不需要血红素而直接与铁结合(2 个铁,而不是 1 个铁),而铁包住氧。这样,可以很牵强地找到蚯蚓血红蛋白和细胞色素的某些功能共同性,但是,它们与结合 RNA 的病毒外壳蛋白确实没有任何功能共同性,虽然它们的结构骨架(四螺旋束)的架构很相似。而且相似性不仅在于其架构(四螺旋束),还在于肽链沿这一架构上的通路,即肽链的折叠模式。图下方的拓扑结构示意图(表示了从顶端看过去的螺旋束外观)标示了所有 3 个蛋白质的肽链折叠模式是相同的。

因此,具有相同折叠模式的蛋白质可能具有完全不同的功能。而蚯蚓血红蛋白[1]和经典的与氧结合的蛋白质肌红蛋白[2](图 14-2)执行完全相同的功能(只是前

①血红蛋白(hemoglobin)是脊椎动物红血细胞中一种含铁的呼吸色素,由大约 6% 的血红素(heme,或译为血色素、亚铁原卟啉、亚铁血红素)和 94% 的珠蛋白(globin)组成。hemoglobin 曾经包含了血红素和血红蛋白两个意思,现在区分开了。

②肌红蛋白(myoglobin)是肌肉内储存氧的蛋白质,是由一条肽链和一个血红素辅基组成的结合蛋白。

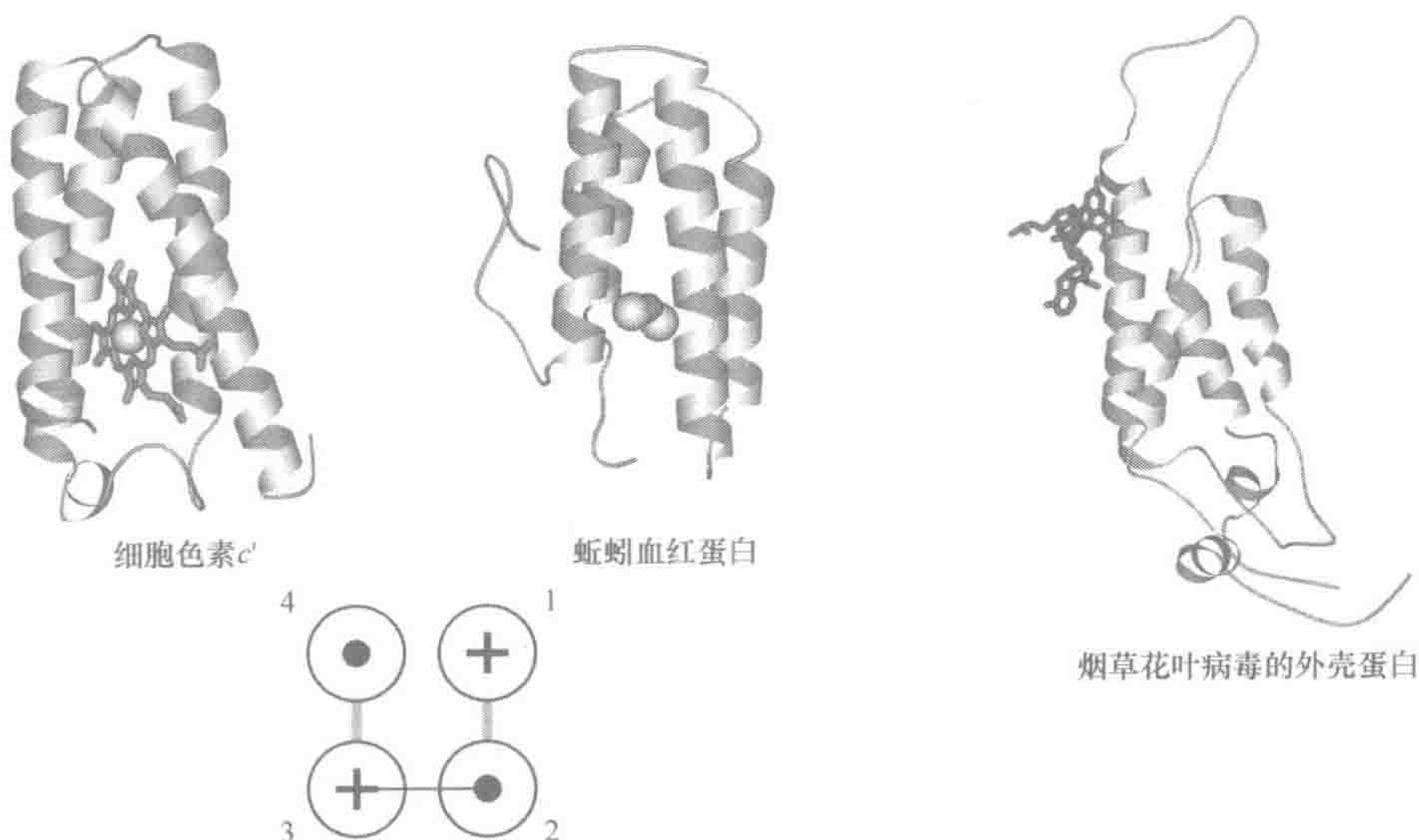


图 14-1　3 种架构相似（"四螺旋束"）、但功能不同的 α 螺旋蛋白质：细胞色素 c'、蚯蚓血红蛋白（参见彩色立体图对 D-21）和烟草花叶病毒外壳蛋白。图中同时显示了蛋白质链和辅因子：骨架模型——血红素（细胞色素中）和 RNA 片段（烟草病毒中），小球——铁离子（在细胞色素的血红素中和蚯蚓血红蛋白中），以及与铁结合的氧（蚯蚓血红蛋白中）。这些螺旋束的总体架构类似于 β 片的纵向拼装。拓扑结构示意图（下方）显示了"从下方"（从结构片段的顶端）观看时的蛋白质外观。圆圈表示为 α 螺旋的末端。十字对应于片段的 N 端（即它"远离我们"），黑点表示它的 C 端（即它"走近我们"）。图中显示了连接结构片段的环线的通路：黑线表示朝向我们的，灰色表示处于拼装包的背面。示意图中的数字表示结构片段在肽链中的顺序（从 N 端到 C 端）

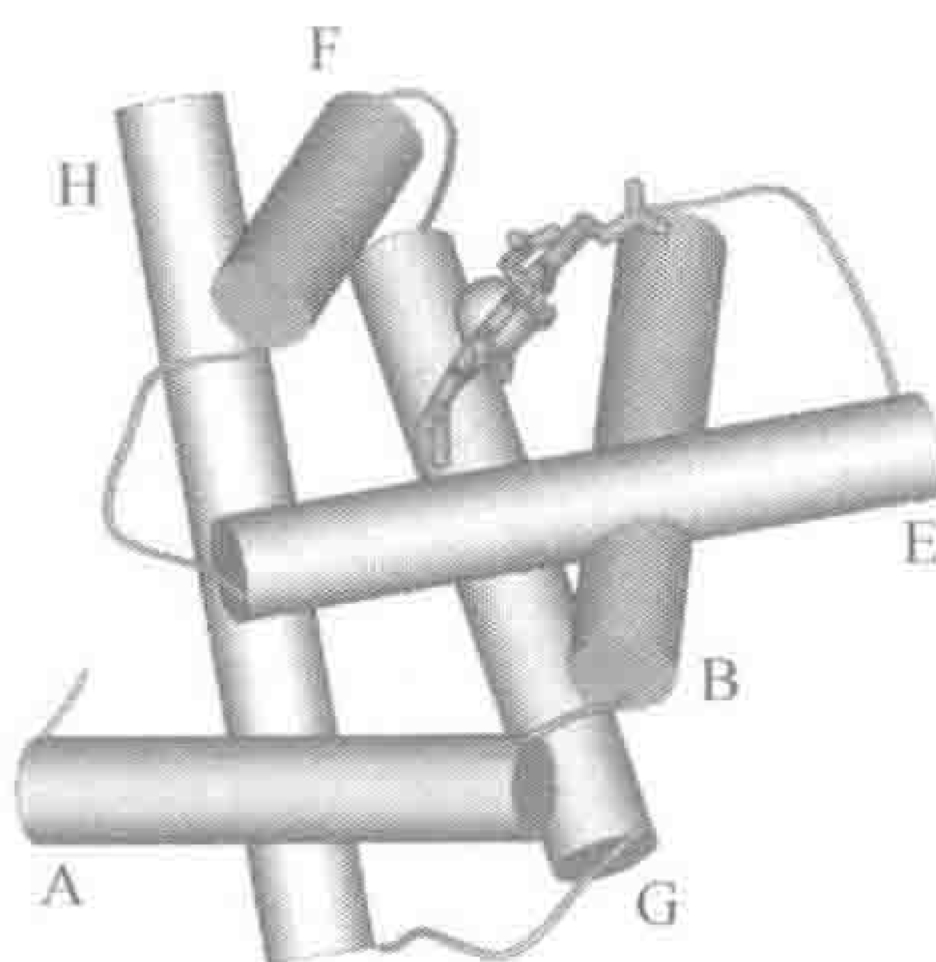


图 14-2　肌红蛋白的结构：各由 3 个 α 螺旋组成的交叉层。螺旋 A、E 和 F（按照在肽链上的顺序而命名）位于上层，螺旋 H、G 和 B 位于下层。短螺旋 C 和 D（每个只有 1～2 圈）未画出，因为它们在肌红蛋白中并不是保守的。血红素位于上层的裂隙处。这种"交叉层"类似于 β 片的正交拼装。（螺旋 B 与 E 的正交接触特别紧密，因为在它们中的甘氨酸所处位置在螺旋表面形成的凹部进入了另一个的表面上的相同的凹部）

者在血液中，而后者在脊索动物中，包括我们人类），但是除了它们都是 α 型蛋白质之外，它们的架构却截然不同。在蚯蚓血红蛋白中所有 α 螺旋都是平行排列的，而在肌红蛋白中 α 螺旋组织成垂直的两层。这又是一个例子，说明相似的功能可

以由架构不同的蛋白质来完成，而相同构造的蛋白质可能执行完全不同的功能。

我想你们都知道，同源的蛋白质（如肌红蛋白和其他珠蛋白）在结构上相似，并且执行同样的功能，所以我，像前面一样，想把你们的注意力引到非平庸的实例上来：在这些例子中，在蛋白质的结构和功能之间的没有什么联系。

通过比较肌红蛋白和蚯蚓血红蛋白，我想请你们注意，这两种蛋白质的活性中心（在前一种中：内部的带有铁离子的血红素，在后一种中：两个铁离子）都位于结构的“架构性缺陷”中，在这些实例中，是在被分开的螺旋之间的缺口处。

■ 上面所讨论的“螺旋束”对于长 α 螺旋是很典型的。它们不仅在水溶性球状蛋白中被观察到，也在纤维蛋白和膜蛋白中被观察到。在这种情形下，被封装在 α 螺旋之间的蛋白质核心（在水溶性球状蛋白和纤维蛋白中是疏水的，在膜蛋白中是亲水的）呈伸长的准圆筒形。“交叉层”也是由相当长的螺旋所形成的；这种拼装包具有扁平的疏水核心。

■ 然而，对于在球状蛋白中典型的相对较短的螺旋（它们的长度通常为大约 20Å，直径 10Å），更典型的是螺旋围绕由它们的侧链基团所构成的“准球状”疏水核而产生准球形折叠。

球状蛋白中的螺旋的典型拼装显示于图 14-3 中。它们不能以螺旋的纵向拼装和正交拼装这些术语来描述：螺旋间的接触角度绝大部分约为 40°～60°。

然而如此复杂的折叠可以借助于“准球形多面体”模型很好地描述并分类，该模型是由 A. G. Murzin 和 A. V. Finkelstein 在 20 多年前提出来的。例如，图 14-4 显示了这种模型如何描述在图 14-3 中所示的 α 螺旋球。顺便说一句，这两幅图片取自 *Nature* 杂志上介绍上面所讲的 A. G. Murzin 和 A. V. Finkelstein 的工作的一篇综述文章。

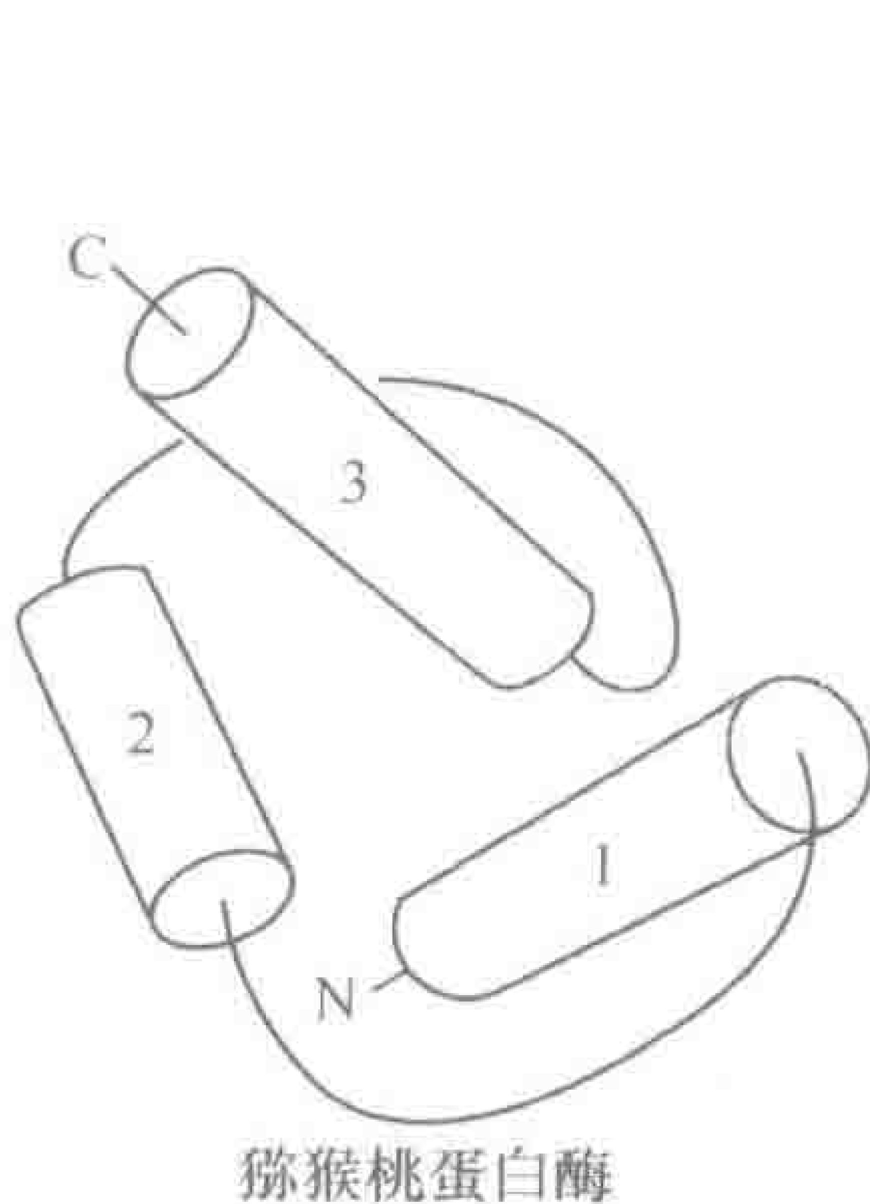


图 14-3 球状蛋白中的典型拼装：猕猴桃蛋白酶的 N 端域（环线的通路描绘得非常粗略）。注意，该域的架构不能用 α 螺旋的纵向拼装和正交拼装等术语来描述

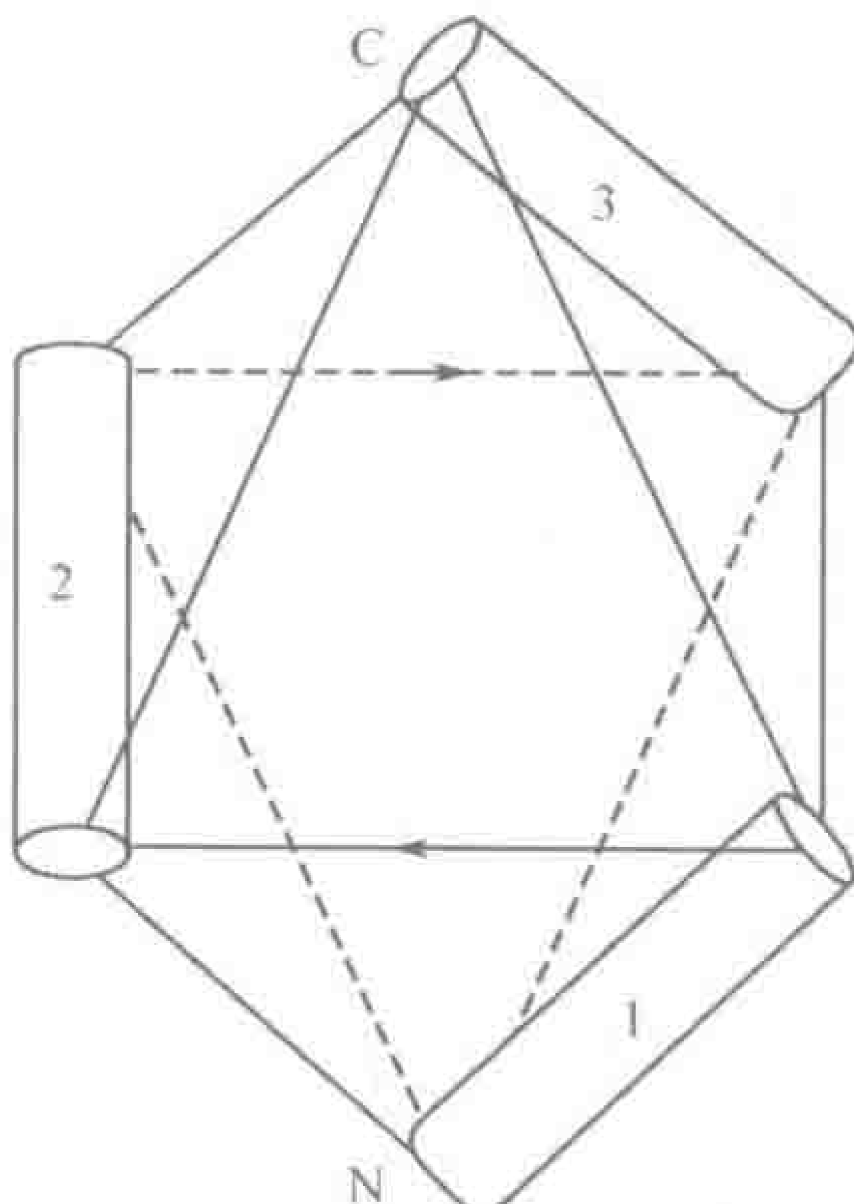


图 14-4 α 螺旋在准球形多面体各边上的排布，以模拟猕猴桃蛋白酶的 N 端域。箭头表示连接 α 螺旋的环线的通路

我不能否认展示来自同一综述文章中的另外两幅图片(图 14-5)的喜悦心情。

顺便说一句,准球形多面体模型也能相当合适地描述相当长的螺旋束——我们在图 14-1 中所看到的那些。这显示在图 14-5b 中。

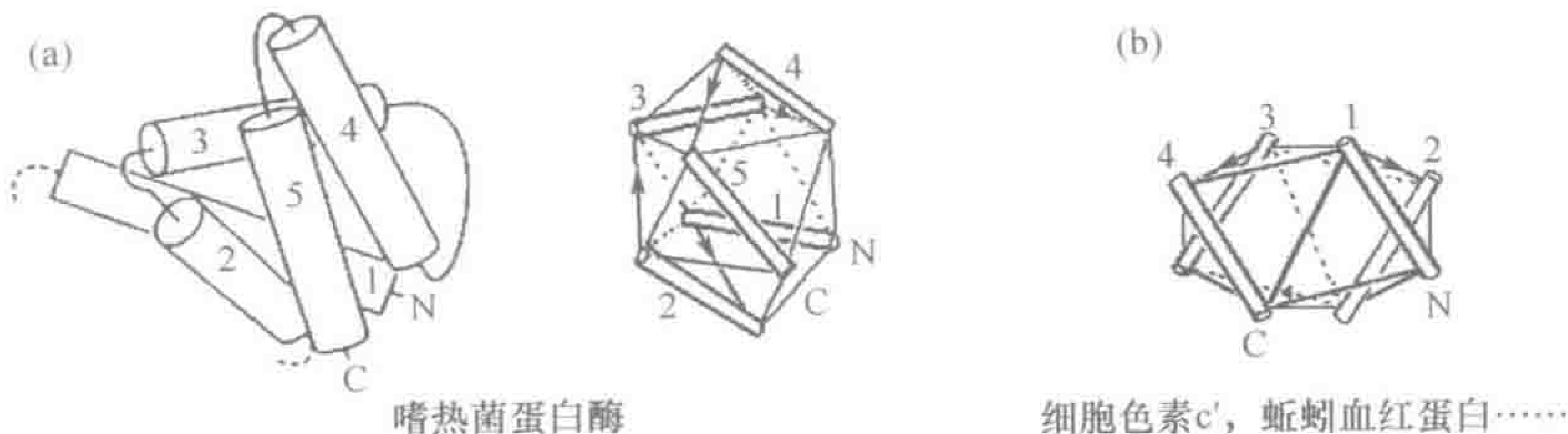


图 14-5 用准球形多面体模型描述螺旋拼装成球状蛋白的其他例子。(a)嗜热菌蛋白酶的 C 端域和基于螺旋在准球形多面体各棱边上排布的模型;(b)图 14-1 中的四螺旋球状蛋白的模型。图 14-5(a)(以及图 14-3,图 14-4)取自上文提及的综述(Chothia C. 1989.Nature. 337:204-205),略有修改

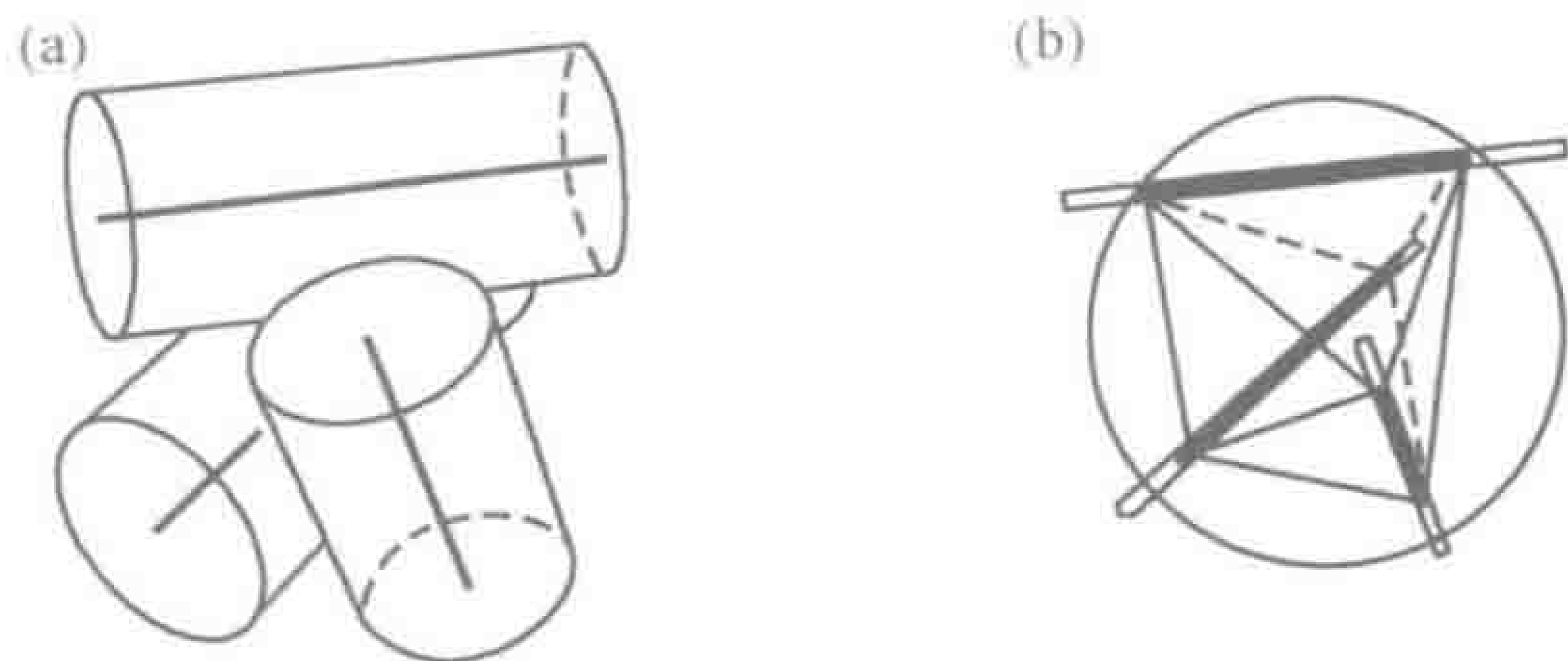


图 14-6 用于描述螺旋拼装的多面体的构造。(a)三条 α 螺旋(带轴的圆筒,直径 10Å)的紧凑拼装;(b)多面体的构造:用半径 10Å 的球壳包围拼装中心,球壳与螺旋轴的交叉处给出了多面体的顶点。螺旋轴排布在它的三条棱边上。棱边在球内的部分显示为黑色。轴与球壳的交叉点形成多面体的顶点。每个顶点对应于一个螺旋的一端。螺旋轴形成多面体的一部分棱边,而其余棱边描绘了螺旋的接触

准球形多面体模型的本质在于,它把我们的注意力集中到螺旋围绕蛋白质球状核心的排布。模型只考虑 α 螺旋(伸直的坚硬粒子)紧凑包围核心;而螺旋的极性端应该位于球体表面。每一种螺旋拼装都可以用多面体来模拟(图 14-6),其每一个顶点都对应着螺旋的一端。紧密的球状蛋白可由最紧凑的"准球形"多面体(图 14-7)来描述。在球状蛋白中观察到的拼装包都接近于理想拼装。每一定数目的螺旋对应着一个多面体,在此框架内存在有几种(从 2 至 10)折叠类型,对应于螺旋轴在该多面体的棱边上的不同排布方式。在这些折叠中包括了上面讨论过的"螺旋束"和"交叉层"。

■ 有趣的是,在所观察到的 α 螺旋蛋白质的架构中,沿准球形多面体各棱边走向的不仅有螺旋,通常还有连接螺旋的不规则环线(图 14-4,图 14-5)。换句话说,在典型的情况下,蛋白质链通过沿准球形多面体的棱边上的连续路径封装它的疏水核心。

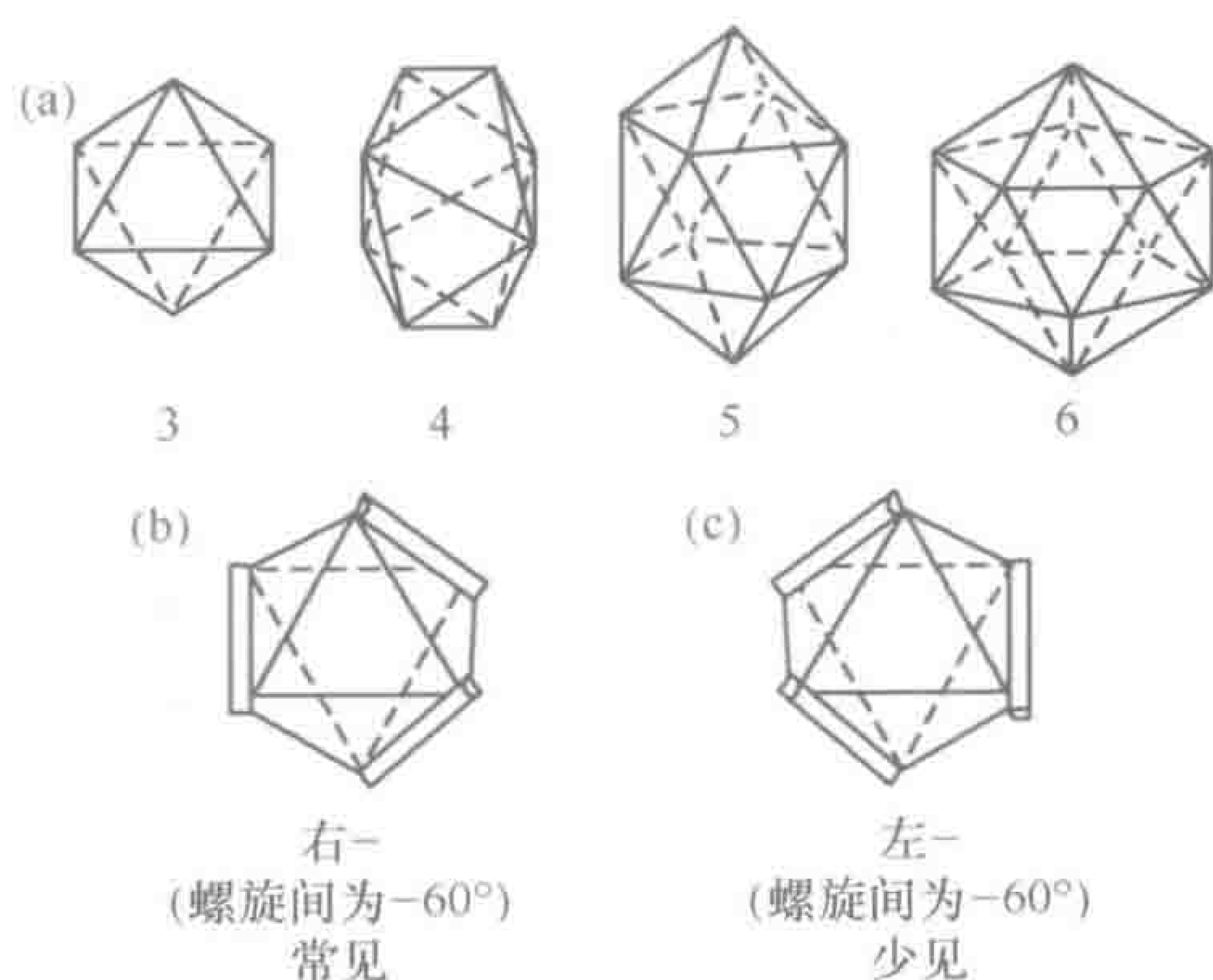


图 14-7 分别描述 3、4、5、6 条螺旋的紧凑拼装的准球状多面体(a)。数目更多的螺旋不能围绕球状核心而拼装。每一个多面体描述了几种折叠模式,即对应于螺旋轴在其棱边上的不同排布的螺旋“堆码”类型。三螺旋复合体有 2 种折叠类型[(b)右手型螺旋束;(c)左手型螺旋束(图 14-3,图 14-4)],四螺旋的有 10 个,五螺旋的有 10 个,六螺旋的有 8 个(对于四至六螺旋的球状蛋白的“堆码”未显示于图中,但是如果希望看到的话,只要以所有可能的使多面体的每个顶点对应于一个螺旋末端的方式把螺旋排布在多面体的棱边上,就可以轻易地自己构造出来)。螺旋间的角度有助于螺旋紧密接触的那些拼装包(图 14-9),比其余的在蛋白质中更经常被观察到

■ 现在来看一个问题,蛋白质球中的紧凑拼装是如何形成的。实验显示,蛋白质象有机晶体那样紧密和坚硬,证实了这种拼装的存在。然而仍然有待阐明的是,这样的拼装是如何达成的,因为蛋白质链的形状实在太复杂了,侧链基团也非常种类繁多。

更准确点,紧凑拼装的形成原则,在最笼统的意义上,仅就 α 螺旋而言是多少比较清楚的,所以现在正好是讨论这个问题的时候。

α 螺旋紧凑拼装的第一个模型,是按照“凸起(侧链基团)进凹处(侧链基团之间)”的原理进行拼装的,它是由 F. Crick 于 1953 年提出的,那时还没有解析出任何一个蛋白质的三维结构。然后它被 A. V. Efimov 和 Chothia-Levitt-Richardson 研究组分别独立地发展,现在获得了“脊骨(侧链基团)进凹槽(侧链基团之间)”的模型形式。

按照这一模型,α 螺旋表面的侧链基团形成凸起,形成由凹槽所分隔的脊骨。“脊骨和凹槽”比“凸起和凹处”能稍微更好地描述实际情况,因为一个凸起(一个侧链基团)朝另一个凸起(另一个侧链基团)的盘旋可使这个或那个“由凸起形成的脊骨”更加确切。脊骨(和沿着它的凹槽)有两种类型。“+4”型脊骨由分布在链上编号为“i”、“i+4”、“i+8”(或者说,周期为 4)等残基的侧链基团构成,“+3”型脊骨由编号为“i”、“i+3”、“i+6”(即周期为 3)等残基的侧链基团构成。图 14-8 显示,这些脊骨与螺旋轴形成了正负号不同的倾角。

在紧凑拼装中一个螺旋的脊骨进入另一个螺旋的凹槽。这样就有两种基本的可能性(图 14-9)。

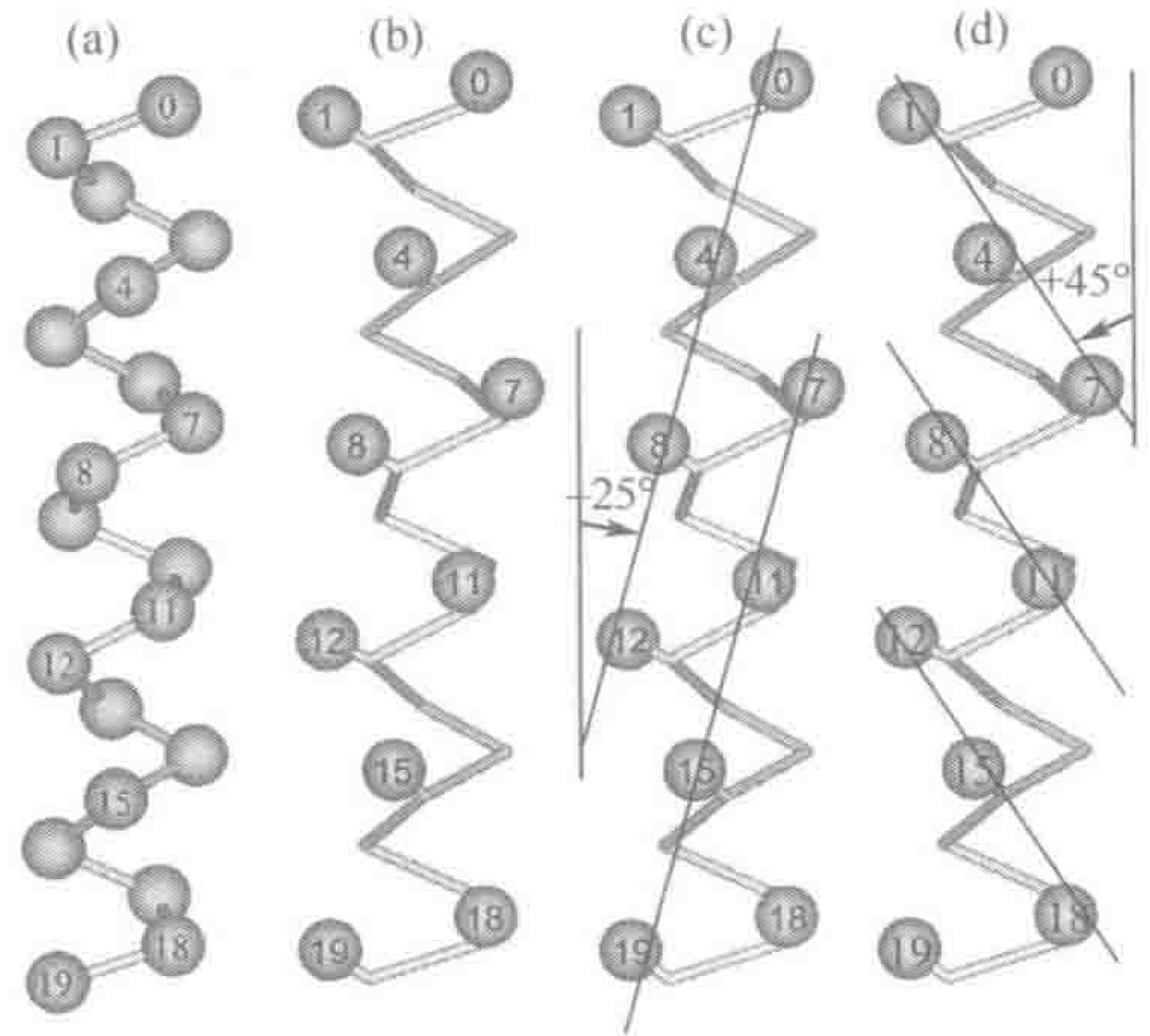


图 14-8 α 螺旋;标出了 C^{α}(a)和 C^{β}[(b)~(d)]。编号的残基朝向读者。显示了两种类型的由相邻侧链基团构成的脊骨(螺旋正面的粗线)[(c),(d)]。由侧链基团“i”-“$i+4$”-“$i+8$”…构成的脊骨沿与螺旋轴呈-25°走向(c),由侧链基团“i”-“$i+3$”-“$i+6$”…构成的脊骨沿与螺旋轴呈+45°走向(d);角度在图中显示较小,因为典型的脊骨是穿过硕大的侧链基团的,而在图(c),图(d)中是穿过 C^{β} 原子的中心。图片取自文献[5],略有修改

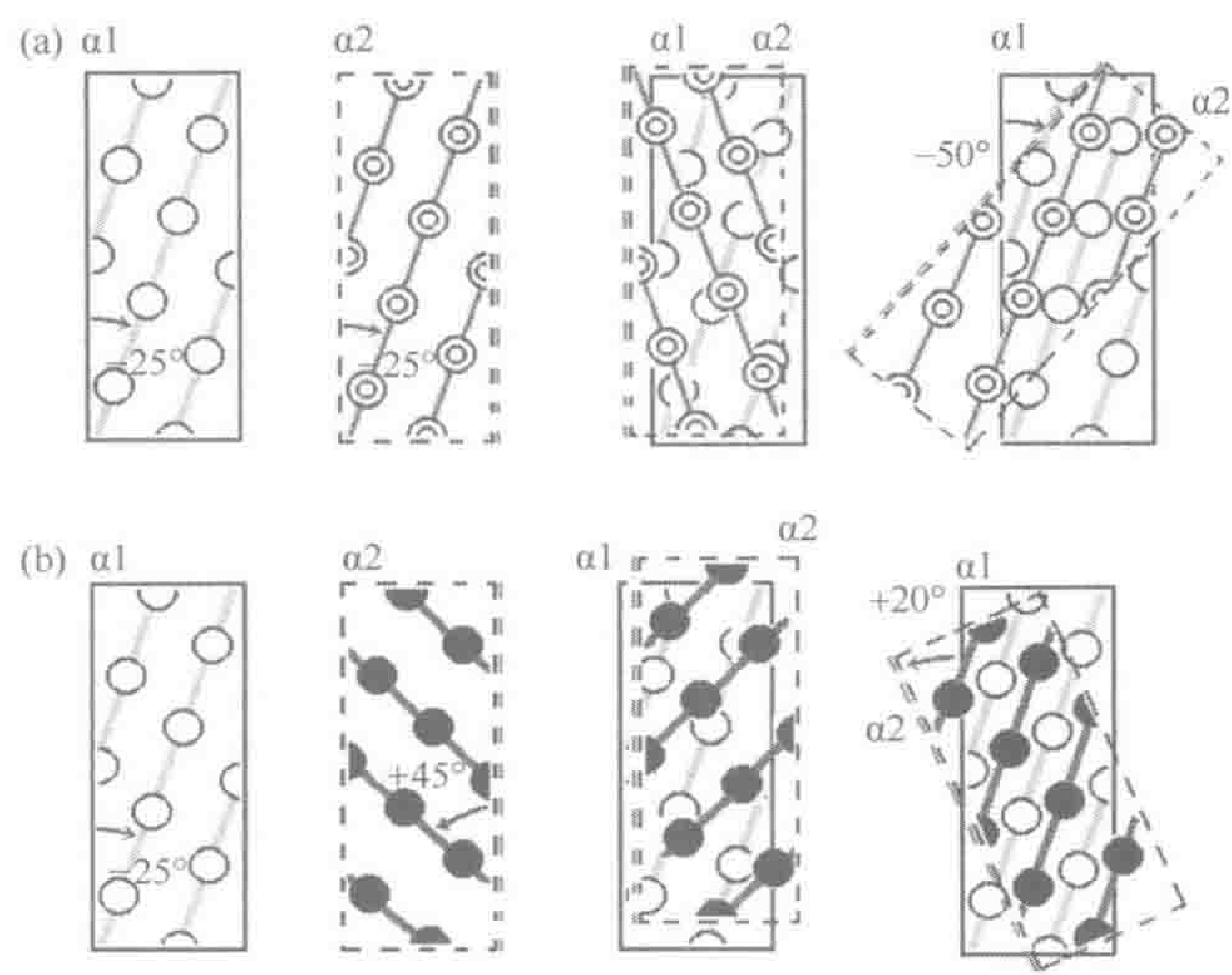


图 14-9 在螺旋接触时侧链基团紧凑拼装的两种基本方式:-50°角(a)和+20°角(b)。我们看到的是沿一条螺旋的接触带(透过沿自身轴翻开的 α2)。“下面的”螺旋 α1 的残基显示为浅色小圆圈,而上面的螺旋 α2 的残基显示为黑色小圆圈。图片取自文献[5],略有修改

(1) 一个螺旋的“+4”型脊骨嵌入到另一个螺旋的同样的“+4”型脊骨之间的凹槽[图 14-9(a);可见,为了获得紧凑拼装,螺旋 α2 倒转过来,垒放在螺旋 α1 上并扭转,直到两个螺旋的“+4”型脊骨相互平行为止]。在这种拼装中螺旋轴之间

的夹角接近于−50°。这个角度是α螺旋球状蛋白中的螺旋接触最大的特征。它也对我们下面要讲到的α/β和α+β型蛋白中的螺旋接触也很典型。重要的是，在这种螺旋接触下α螺旋层的扭转度（扭转角约为−50°/10Å，其中−50°为相邻螺旋的轴间夹角，10Å为α螺旋的宽度）与β片的典型扭转度（具有同样的扭转角，−25°/5Å，其中−25°为相邻β串的轴间夹角，5Å为β串的宽度）吻合得不错。

（2）一个螺旋的"+3"型脊骨嵌入到另一个螺旋的"+4"型脊骨之间的凹槽中［图14-9（b）］。在这种拼装下，螺旋轴间的夹角约为+20°。这个角度对于螺旋束中的螺旋接触最为典型，在α螺旋球状蛋白中、纤维蛋白中和膜蛋白中也是如此。

除此之外，一个螺旋的"+3"型脊骨也可以嵌入到另一个螺旋的同样的"+3"型脊骨的凹槽中，形成几乎垂直正交的螺旋的极短接触。由于这种接触的尺寸很小，它在图14-9中未显示出来，虽然螺旋的这种"垂直正交"接触对α螺旋球状蛋白也相当典型。

在总结对紧凑拼装的描述时，应该注意到，与上述的"理想的"角度的实际偏离非常大，因为侧链基团的尺寸差别很大。根据同样的原因，在β结构中（其中侧链基团突出较小，因而β片的表面相当平坦，而螺旋则是凸起不平的）脊骨嵌入到凹槽中的图景非常模糊，只在个别的情形下能观察到。

■　最后。螺旋的紧凑拼装与前面讲到的螺旋在准球形多面体的棱边上排布模型符合得怎么样呢？原来，是以相当有趣的方式。那些螺旋之间的角度接近于螺旋紧凑拼装所需的−50°和/或+20°的"多边形"拼装，经常碰到；而其他的则罕见（但也能观察到）。例如，在图14-7中所示的两种三螺旋拼装中，其中一个**左**旋的螺旋束使螺旋间的夹角为−60°［接近于紧凑拼装所需的−50°，图14-9（a）］，这种三螺旋束经常碰到；而另一个**右**旋的螺旋束使螺旋间的角度为+60°（它远离对紧凑拼装最佳的所有角度：−50°、+20°和90°），这种三螺旋束的出现频度低一个数量级。

■　现在来看由β片和α螺旋组成的"混合"蛋白质。分层结构是它们的特征，而且α螺旋和β段不能位于同一片层上，这会导致β片边缘的氢键出现能量不利的脱水（图14-10）。



图14-10　混合型（α/β和α+β）蛋白质的分层结构。从α螺旋（为了强调球状蛋白拼装的紧凑性，末端以正方形表示）和β串（从顶端观，长方形）顶端观。α螺旋和β串不能处于同一片层上，这会导致β片边缘处的氢键脱水（β片中氢键的供体和受体用黑点表示）

"混合型"蛋白质可以分为α/β（"α杠β"）型和α+β（"α加β"）型蛋白质（更确切些：域）。有时它们组合成常见的α&β（即"α和β"）型，或者α−β型蛋白质。

在α/β型的域中，β结构是平行的，α螺旋也彼此平行（且反平行于β段），而

α 和 β 段在肽链中的特征交替具有如下形式:α-β-α-β-α-…。

α/β 型蛋白质有两种最典型的构造模式:α/β 圆筒,其中 β 圆筒处于由 α 螺旋所组成的圆筒的内部[图 14-11(a),参见彩色立体图对 D-7 和 D-27],和“Rossmann 折叠①”,其中多少有点平坦的[除了普通的(即右旋的,如果顺着 β 串看的话)螺旋桨式扭转之外]β 层处于与之互补的扭转的 α 螺旋层之间[图 14-11(b),参见彩色立体图对 D-26]。与前面所讨论的域不同,α/β 域通常具有两个疏水核心:在 Rossmann 折叠中是处于 β 片和各螺旋层之间;在 α/β 圆筒中是在 β 圆筒内部(较小的核心)和在 β 圆筒与 α 圆筒之间(较大的核心)。

顺便说一句,还存在有一种“α/β 准圆筒”,它是“不完整的”,未封闭成圆形(这意味着,在 β 圆筒内部不存在疏水核)。它们被称为“α/β 马蹄铁”,其中共有一个半的 α/β 十倍重复体。

然而,通常在 α/β 圆筒中共有 8 个 α 段和 8 个 β 段,而且几乎所有 α/β 圆筒都具有同样的拓扑结构:所有在肽链上相邻的或相隔一个分布的 β 段和 α 段都彼此接触。看来,这种构造保证了蛋白质球体的异常稳定性,因为大量具有这种架构的蛋白质球(在 10% 的蛋白质中被观察到)在形状上都彼此非常相似,而且常常缺乏在氨基酸序列上共同起源的任何痕迹,缺少功能上的任何共同性。

没有功能共同性?是的;没有活性中心构造上的共同性?是的;但活性中心在蛋白质球架构上的位置并非没有共同性,每一个架构都含有一个好像是这个架构专门为活性中心而创建的位置(通常是凹陷),无论它行使什么功能。

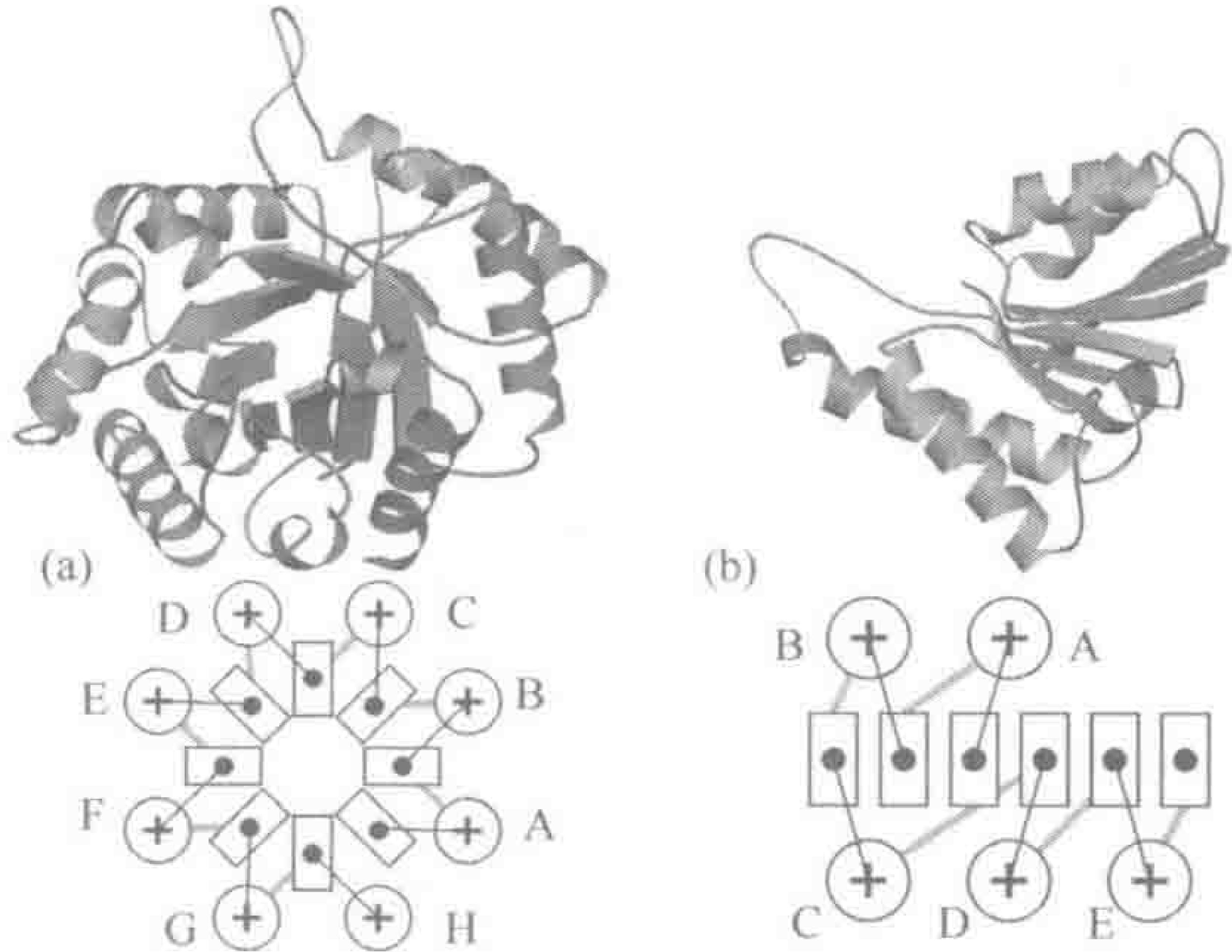


图 14-11　α/β 型蛋白的典型构造模式及其简化模型(模型的视角:从 β 层的末端):(a)磷酸丙糖异构酶中的“α/β 圆筒”;(b)苹果酸脱氢酶的 NAD 结合域中的“Rossmann 折叠”。在细节图上,在第一个蛋白质上可见朝向我们的由发散呈花结状的环线所形成的凹陷,它通向 β 圆筒的中心。在第二个蛋白质中,左上方可见位于下方和上方的环线之间的裂隙

①Rossmann 折叠(Rossmann fold)是 Michael Rossmann 最早(1973 年)指出的在核苷酸结合蛋白经常出现的一种折叠模式。该结构由被 2 个 α 螺旋连接的 3 个或更多的平行 β 串以拓扑顺序 β-α-β-α-β 所组成。

■ 我想请你们把注意力转到 α/β 圆筒轴上的“喇叭口”上来[图 14-12(a)]:可以看到,这个凹陷摆放在蛋白质的总体架构中,就在它的肽链折叠模域(motif)中,未被环线所覆盖。活性中心就位于该处。更确切一点:在分布于 α/β 圆筒两个相对端面上的两个这样的“喇叭口”中,仅有一个被用于容纳活性中心,即 β 段 C 端和 α 螺旋的 N 端所指向的那一个。有人认为,正是这些末端(有可能,因为螺旋 N 端有大量张开的 NH 基团)以及连接它们的短环线,对于各种基质的结合特别有用。不过,这个问题还不完全清楚。

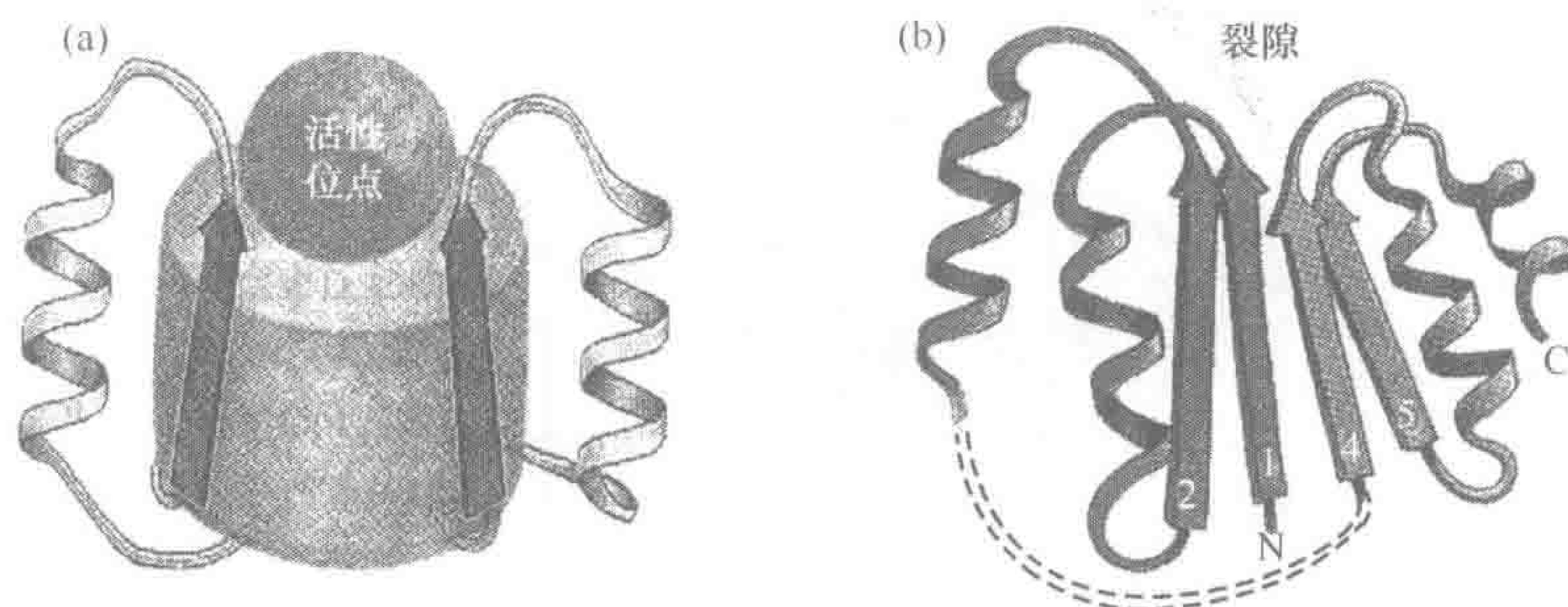


图 14-12 α/β 型蛋白质中的活性中心的典型位置:在 α/β 圆筒轴上的“喇叭口”中(a),在“Rossmann 折叠”中的发散环线所形成的裂隙中(b)。图片经允许取自文献[5]

在“Rossmann 折叠”中观察到活性中心处于相似的位置:裂隙和凹陷处,而且在凹陷处也显露出 β 段的 C 端和 α 螺旋的 N 端。只是这里的凹陷不是由从圆筒中心所发出的环线所形成,这些环线一部分是从 β 片通向处于其下方的 α 螺旋,另一部分是从 β 片通向处于其上方的 α 螺旋。

课堂里的声音:只是不应该认为,活性中心总是位于这样明显凹陷处。这很常见,但远非总是如此!

讲课人:说得对。通常(在约 80% 的情形下),活性中心位于球状蛋白上最大的凹陷处,而凹陷,通常是由该球状蛋白以二级结构形成骨架的架构所决定的。但是很多情况下,即使已知蛋白质的整个空间结构,寻找其活性中心也总是要花费很长时间,而且并非总能成功地找到……

■ 现在来看 α+β 型蛋白质。它们是以反平行(而非平行的,如 α/β 型蛋白质中那样)β 结构为基础的。

α+β 型蛋白质中可以分为两类。一类[有时被称为“αβ 褶皱(αβ-plaits)”]类似于 α/β 型蛋白质,其中的 α 螺旋层位于 β 片之上。它们也像 α/β 型蛋白质那样,在肽链上和空间上 α 段和 β 段有规则的交替(但与 α/β 型蛋白质中不同)。另一类(“普通”α+β 型蛋白质)不具有这种交替,在其肽链上 α 结构与 β 结构是“隔开的”。

在 αβ 褶皱的肽链中 α 段与 β 段的典型交替的形式为…α-β-β-α-β…或…α-β-β-β-β-α-β-β…(图 14-13,参见彩色立体图对 D-28)。这里单个的 α 螺旋

位于由偶数个 β 串构成的 β 发夹或 β 片之间。在 αβ 褶皱中沿肽链相邻的 β 段形成**反**平行的 β 发夹；而由于 α 螺旋间的 β 串为偶数个（而不是在 α/β 型蛋白质中那样为**奇**数个），并且由于这些 α 螺旋串的基本共轴性，α 螺旋也形成**反**平行的发夹。有趣的是，蛋白质的这种（或者，更确切些，相似的 αββαββ）“褶皱的”结构首先由理论预言到（更确切些，是作为蛋白质工程设计的计划而提出），然后这种架构在自然界中被广泛观察到，而且它们尤其大量地存在于铁氧化还原蛋白中和……RNA 结合蛋白中。

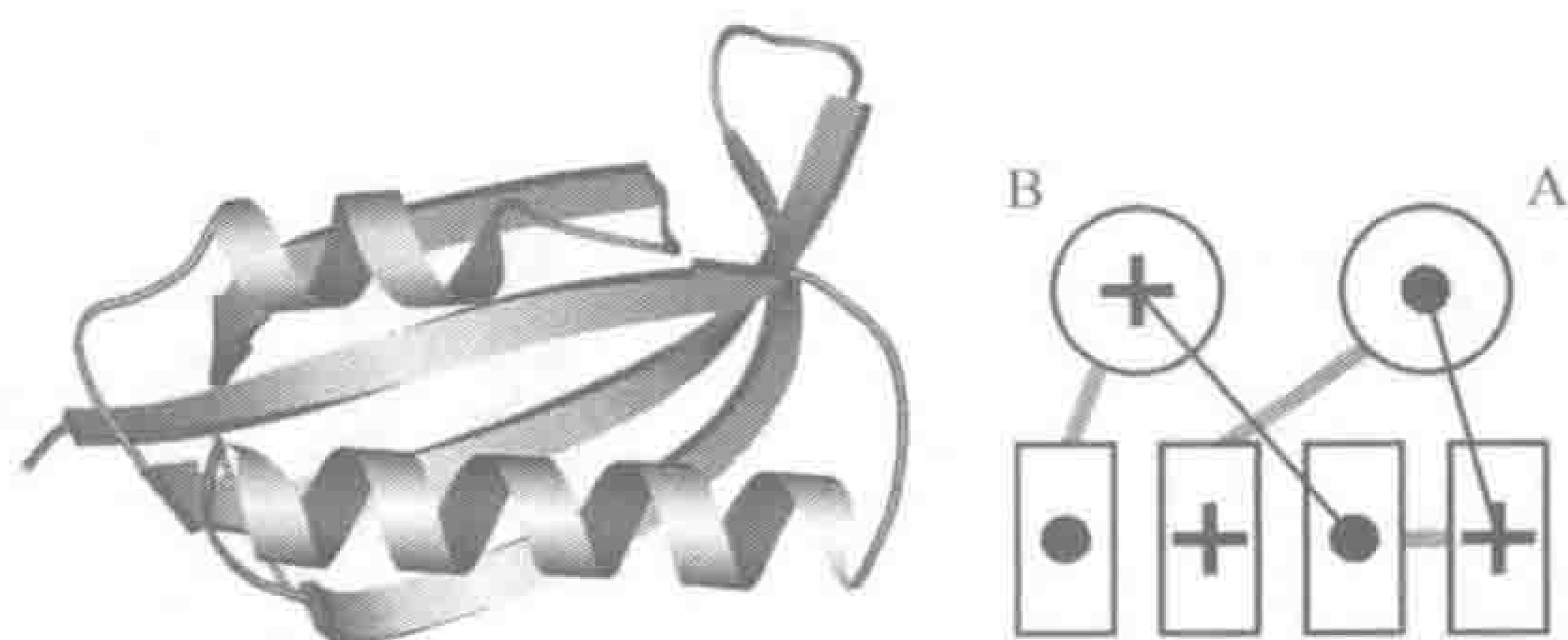


图 14-13　α+β 型蛋白质的一种典型构造模式：核糖体蛋白质 S6 中的“αβ 褶皱”。αβ 褶皱比“普通”α+β 型蛋白质在链中具有更规则的二级结构交替（在图中为 βαββαβ，这就是所谓“铁氧化还原蛋白折叠”）。右边为该蛋白质的构造示意图（从近似于圆筒的结构片段的末端看过去）。螺旋以字母编号。α 段和 β 段末端的“+”表示它远离我们（即其 N 端朝向我们），点表示走向我们

在“普通”α+β 型的域中（图 14-14），α 段和 β 段在肽链上的排列是不规则的，且倾向于形成模块。这类蛋白质通常看上去像被分立的 α 螺旋或 α 螺旋子域所遮盖的 β 片（它常常自身弯曲而形成子域）；在 α+β 域中的 β 结构基本上是反平行的，与“纯”β 型蛋白质中一样。

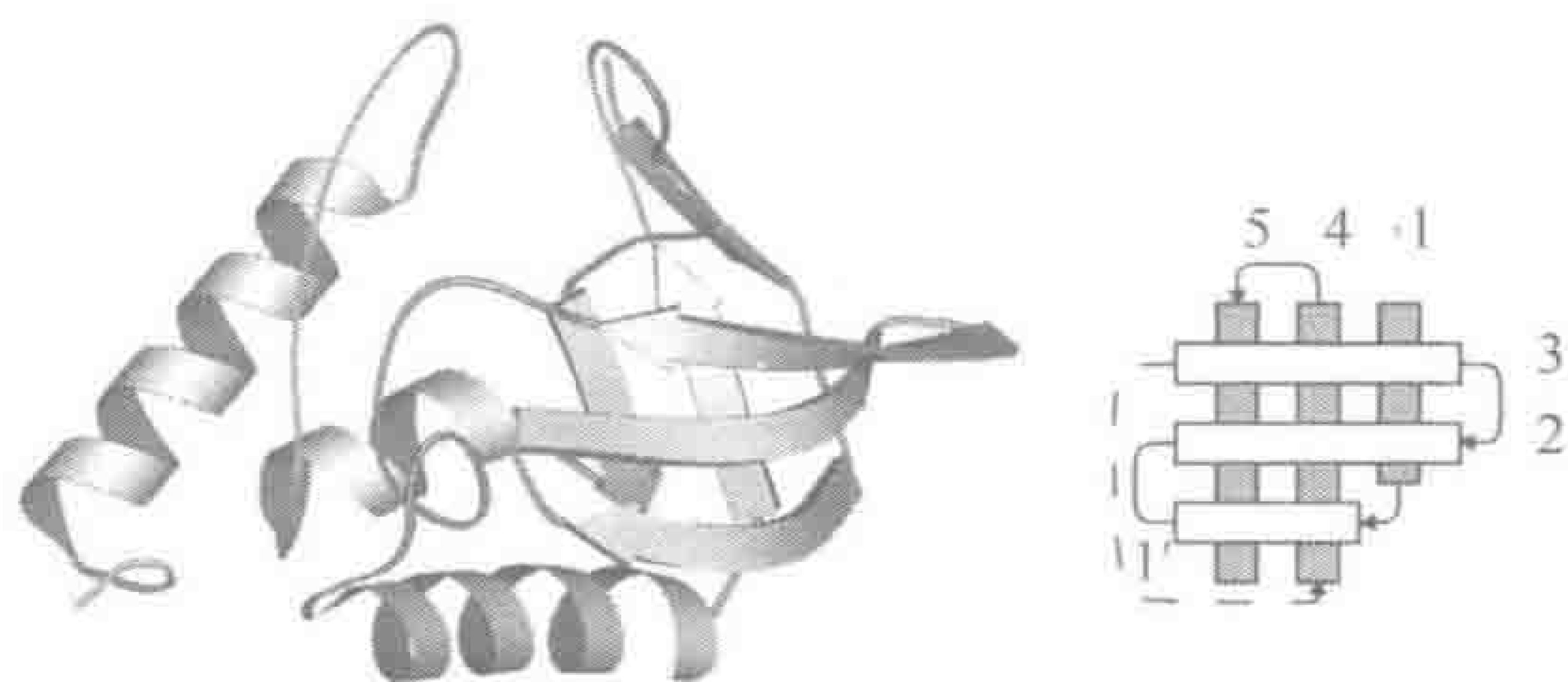
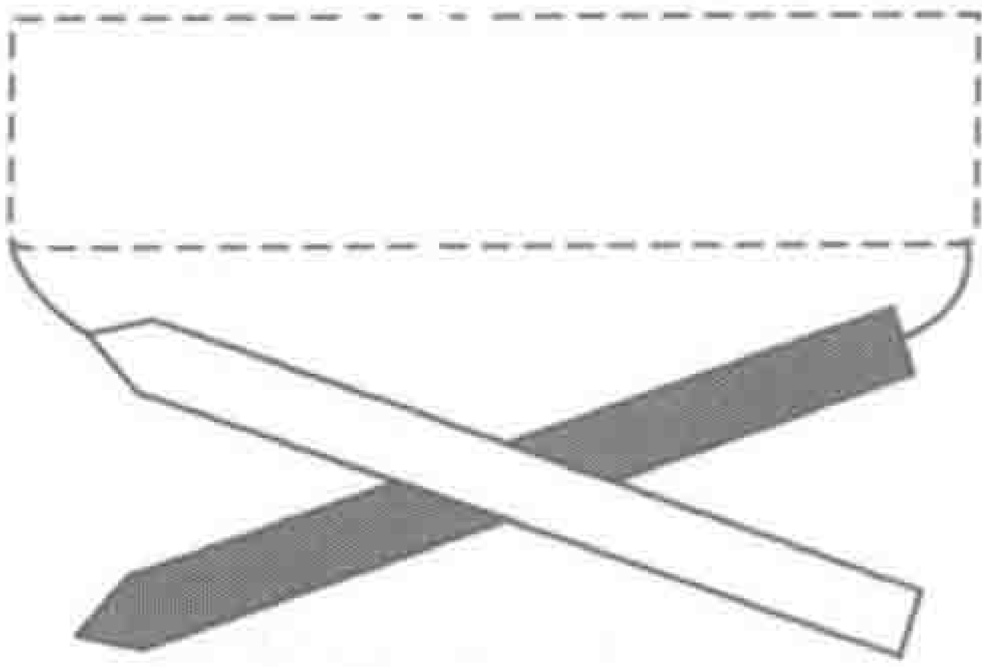


图 14-14　α+β 型蛋白质的典型构造模式：葡萄球菌核酸酶。“普通”α+β 型蛋白质的肽链上二级结构交替（图中为 βββαββαα）不如 α/β 型蛋白质或 αβ 褶皱那么规则。在核酸酶的 β 域中观察到的肽链折叠模式，被称为“OB 折叠”（“低聚核苷酸结合折叠”，即“oligonucleotide-binding fold”，参见彩色立体图对 D-23）。右边为这种经常在不同的多域和单域蛋白质中碰到的 OB 域的 β 结构构造示意图（β 段的正交拼装的顶部观）。β 串以数字编号。第一个 β 串是折弯的（实质上，是断裂的）；它的两部分编号为 1 和 1′。注意“套娃效应①”：一种典型的结构（OB 折叠）是另一个同样典型的结构（核酸酶折叠）的一部分

①套娃：俄罗斯特产木制玩具，一般由类圆柱形的空心木娃娃一个套一个组成，最多可达十多个。

■　α/β 型和 α+β 型蛋白质(以及 β 型蛋白质)的一个非常典型的特点是在肽链上彼此相继、处于同一个 β 片中(但并不一定在这个 β 片上相邻)的平行 β 段(图 14-11 及其他)之间的连接线通路的***右***手螺旋性(即“以逆时针旋转方向接近我们”)(图 14-15)。而且 α/β 型和 α+β 型蛋白质中在这些平行 β 段之间的连接线中通常包含有一个 α 螺旋。在 β 型(有时在 α+β 型)蛋白质中在这种右手螺旋的连接线中,在一个片上的平行 β 段之间包含有[你们应该在上一讲中记得(“abcd”结构等)]另一个片的 β 段,有时甚至单独的 β 片。有时(相当罕见)在平行 β 段之间的连接线中既没有 α 结构,也没有 β 结构,但该连接线通常是***右***手螺旋的,而不是左手螺旋的。


图 14-15　一个片中的平行 β 串之间的连接线的典型***右***手螺旋通路。在连接线中常常还存在有另一个二级结构段

在下一讲中我们将看到,连接线的这种通路通常有助于蛋白质的稳定性,而且正是它允许存在种类繁多的能编码该结构(即稳定该结构)的序列,因此这种通路更经常地被观察到。

■　在粗略地介绍了球状蛋白质的结构之后,我想再次强调,同样的(或非常相似的)架构经常在很多在功能和种系发生上都完全不同的蛋白质中见到。这一发现是对蛋白质进行物理分类,或者理性分类(在文献中更多地这样称呼)的基础。对此我们也将在下一讲中讨论。

第15讲

蛋白质结构的分类。未观察到蛋白质链折叠的“宏观进化”，虽然观察到了它们的“微观进化”。基因复制与专一化。域混合产生的进化。“标准的”三级结构。氨基酸在球状蛋白一级结构中的“准随机”交替的典型性，与纤维蛋白的周期性结构和膜蛋白的块状结构的对比。蛋白质球体构造的物理原理。在蛋白质球体结构中观察到的基本规律：存在单独的 α 和 β 层；罕见有环线的交叠；罕见有沿链相邻结构片段的平行；罕见有左 β-α-β 超螺旋。结构中很少存在“能量缺陷”和“熵缺陷”，这些“缺陷”与能使“缺陷”稳定的蛋白质序列的相对罕见性之间的关系。多数原则。

在这一讲中我们试图解释一个问题：为什么大部分蛋白质都可以归属到一个很小的标准结构集合中？从上一讲中我们应该得到了这么一个印象。

实际上这里有一个“80% ：20%”原则在起作用。这一原则的原始形式是“80%的啤酒是被20%的人所喝掉”。应用于蛋白质则是：“80%的所有蛋白质归属于20%已经观察到的蛋白质球折叠模式”。在前几讲中我也是把重点放在那些典型结构上。

那么，为什么大部分蛋白质都归属于一个很小的标准结构集？而不是全部（像DNA链那样）？这种相似性是在什么结构水平出现的？这些共同的结构有什么意义：记忆蛋白质的共同起源？功能合理性？或是为了满足标准蛋白质结构的一般性折叠原则所必需的？具有不同起源和功能的蛋白质的相似性是在什么结构水平出现的？

在本讲中我们在定性的水平上讨论这些问题，然后，在积累了关于蛋白质自动折叠的知识之后，我们将更严格地回答这些问题。

■ 在所解析的蛋白质数量很少的时候（大约是20世纪70年代中期），存在有一种观点，认为每一个三级结构都是独一无二的，即进化上不同的家族的蛋白质彼此互不相似。然而，随着蛋白质分子空间构造的信息不断增长，蛋白质球体构造存在某些“标准设计”的观念更为清晰。新解析出的蛋白质（或者，至少是它们的域）越来越显示出与已知蛋白质的构造的相似性，尽管它们的功能和氨基酸序列全然不同。因此开始产生并发展起来、在前面各讲中不止一次地出现过的观点，认为蛋白质三级结构的相似性的原因**不只**在于蛋白质的进化趋异，也不在于（或不只在于）蛋白质的功能趋同，而只是简单地在于某些物理规则对于折叠的一套限制作用。

至20世纪70年代末,已完全清楚,在两个“传统的”结构层次(蛋白质三级结构和其详细的原子三维结构)之间存在有一个过渡层次:已经被我们称为蛋白质链的“折叠模式(folding pattern)”,它由α块和/或β块在球体中相互位置排列所决定,而且正是在这个层次上出现了与进化和功能都不相关的蛋白质相似性。与详细的原子三维结构不同的是,“折叠模式”简单、甚至漂亮得令人称奇(图15-1)。

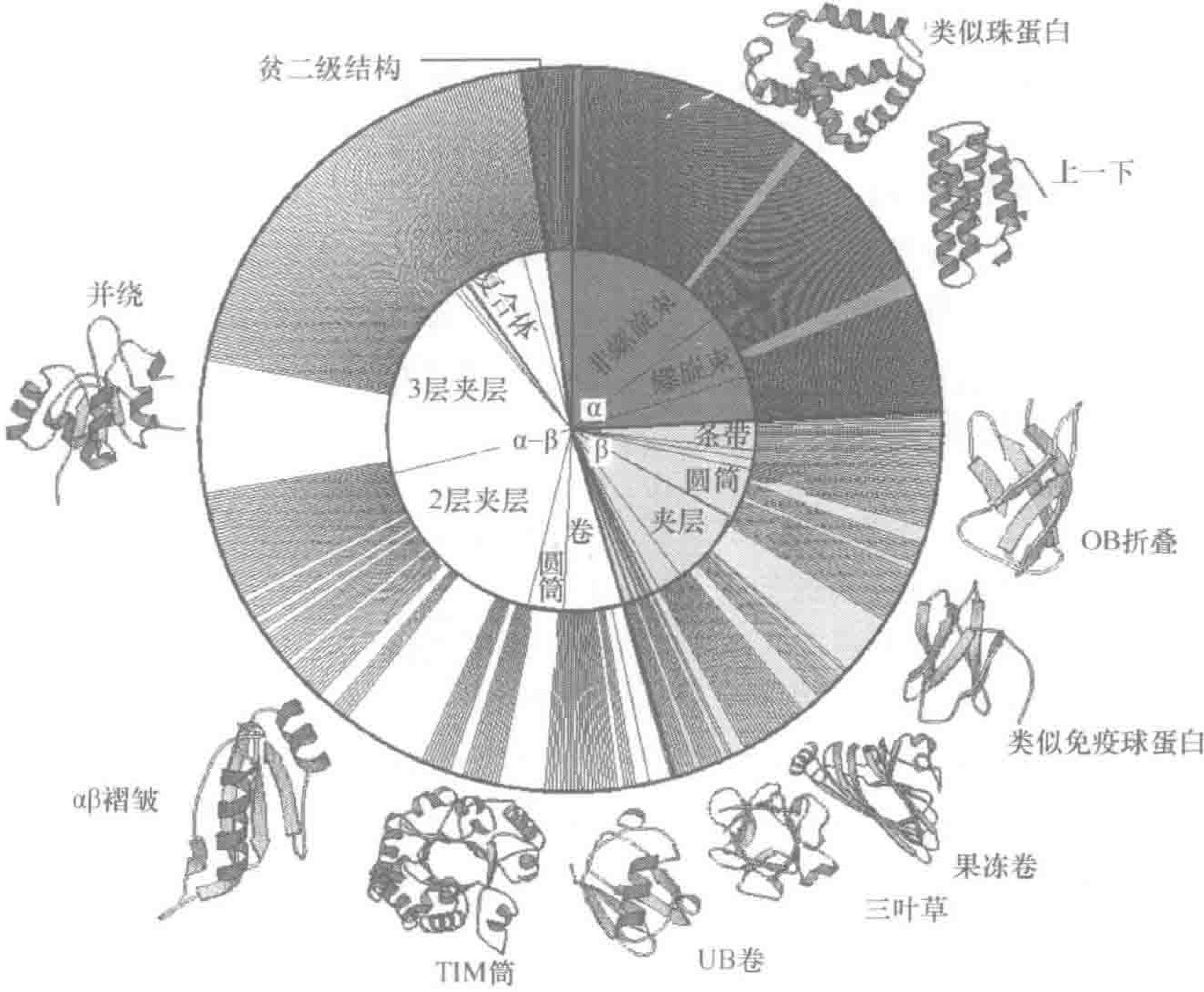


图15-1 蛋白质物理学分类(CATH)下的蛋白质结构类型(“α”、“β”、“α-β”和“贫二级结构①”),典型架构(“非螺旋束”、“螺旋束”、“条带”、“圆筒”、“夹层”、“卷”、“2层夹层”和“3层夹层”以及“复合体”)和典型肽链折叠模式(拓扑结构):类似球蛋白、“上-下”型螺旋束、OB折叠、类似免疫球蛋白、果冻卷、三叶草、UB卷、α/β圆筒,αβ褶皱和并绕。扇形的宽度对应于该类型结构在非同源蛋白中的丰度。注意α结构和α结构的分层拼装中,每一层要么仅由α螺旋构成,要么仅由β串构成,但不会同时由α螺旋和β串构成。图片取自Orengo C. A. , Michie A. D. , Jones S. , et al.1997.Structure.5:1093,略有修改(感谢C. A. Orengo和J. M. Thornton的慷慨许可)

相同的(或相似的)架构经常在功能上和种系发生上完全不相似的蛋白质中碰到。这一发现奠定了对蛋白质进行物理的、或理性的(在文献中更多地这样称呼)分类的基础。

■ 蛋白质结构的最完整分类可由计算机分类程序中实现。这些程序包括:由Holm

①二级结构成分含量很低的蛋白质分子。

和 Sander 研制的"Dali/FSSP①", Thornton 研究组研制的"CATH②"(Class-Architecture-Topology-Homolog),以及(也许是最流行的)"SCOP③"(structural classification of proteins),它是由 A. G. Murzin 在离开普希洛科学中心到剑桥大学之后所研制的。

实际上,分类是针对域(即单独存在的、或作为多域蛋白质的成分而存在的紧密球体)的构造进行的。分类从域的结构**类型**(α、β 等)开始(图 15-1)。结构类型根据由 α 块和/或 β 块构成的骨架的架构(更准确些:架构类型)进行分类,而架构再根据肽链穿行骨架的**拓扑结构**(即按照肽链的折叠模式)进行分类。

更进一步,折叠模式被分成至少显示有某种序列**同源性**(共同起源的痕迹)的超家族,进而被分成具有明显同源性的家族,如此等等,直到具体生物体的单个蛋白质。

蛋白质结构的物理学分类(类型-架构-拓扑结构)不仅能把已知的结构进行分类,而且能预言有待发现的蛋白质结构。例如,长期以来只知道由反平行 β 结构构成的 β 型蛋白质,而由平行 β 结构构成的 β 型蛋白质处显示为空位。然而不久后这一空白就被填补了(还记得 β 棱镜吗?)。

■ 图 15-2 中显示了蛋白质球体的这种结构分类的一个例子,其中为可能的、但(至今?)尚未观察到的肽链折叠模式留有"空位"。有兴趣的是,这一分类(Efimov 树)是基于不同类型的初始小"种子"结构进行逐渐增长和复杂化而建立起来的。这个示意图可以解释为对蛋白质折叠过程(注释一下,完全是由其他实验数据和理论数据得到的该过程的现代观念是,这一过程是从球体折叠形成其一小部分的成核作用开始的,我们将在几讲之后讨论)的模拟。然而这些"树"也可以解释为对蛋白质起源进化的模拟。

■ 有一个总是让生物学家兴奋的问题:我们看到了蛋白质结构的进化了吗?

本质上,这个问题可以分为两个:①我们是否看到了蛋白质的微观进化,即不是简单的"漂移"(蛋白质从生物体到生物体的变化),而是蛋白质以某种形式与所有生物体的变化而相关的那种变化?②我们是否看到蛋白质的"宏观"变化,即它们的结构是否随着生物体结构的复杂化而复杂化?

■ 对前一个问题,我们可以毫不犹豫地回答:是的。虽然远非蛋白质的所有变化都起着明确的功能性作用(木村资生的"中性进化理论"④所强调的),然而在一些情形下蛋白质变化的功能性作用已经得到了很好地理解和研究。例如,羊驼(一种

---

①http://ekhidna. biocenter. helsinki. fi.

②http://www. cathdb. info.

③http://scop. mrc-lmb. cam. ac. uk/scop.

④木村资生(Motoo Kimura,1924-1994),日本生物学家,1968 年提出分子进化中性学说。该学说的主要内容有:在分子水平上,大多数生物演化和物种内的大多数变异,不是由自然选择引起的,而是通过那些对选择呈中性或近中性的突变等位基因的遗传漂变累积引起的。中性学说是达尔文进化论的微观演化水平的进一步发展、修正和补充。

高山动物)的血红蛋白结合氧的能力比住在平原上的亲缘动物强,而且这样的在生存条件作用下的微调通过对成年动物和胎儿的比较可以观察得更清楚:后者应该吸收母体的氧,即与氧结合更强烈。M. Perutz① 指出,在血红蛋白结构上的微观变化是造成结合加强的主要原因。

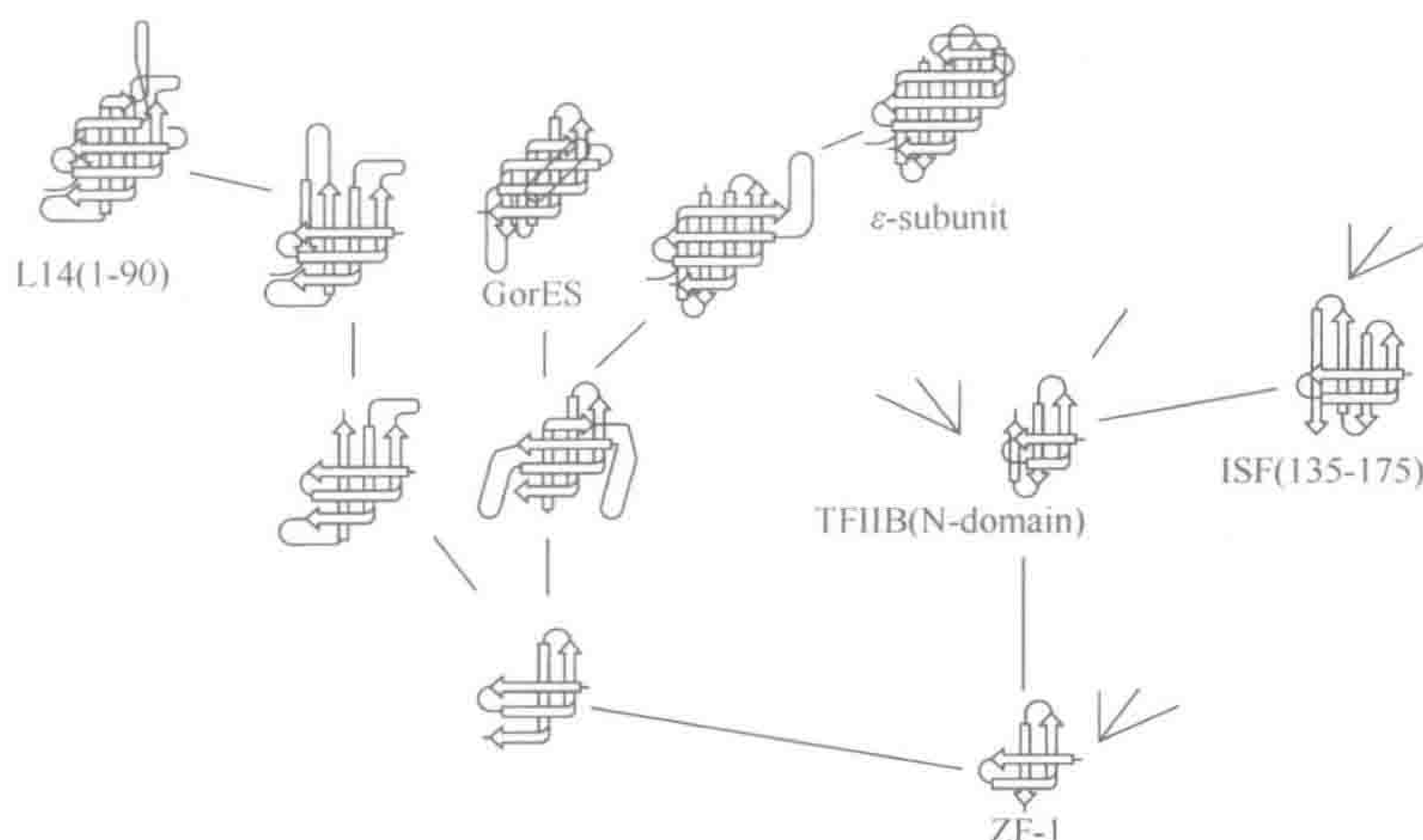


图 15-2 基于初始小"种子"结构的逐渐增长和复杂化的"结构树"(Efimov A. V. 1997.FEBS Letters.407:37-46)的一部分。注意"套娃效应":在较复杂的结构中包含有较简单的结构。显示的树枝片段包含的是一部分基于呈直角弯曲的 β 发夹的 β 型蛋白质。在图中,有名字标记的是在自然界中已经找到的结构,尚未找到的显示为空白

有人认为,进化常常是通过基因的扩增及其拷贝的后续突变而发生的,因此该基因的一份拷贝一直维持"原有的"功能(以及生物体的生命),而其他一份或多份拷贝则能够自由突变,以便(随机地!)寻找能够使蛋白质的功能适应生物需求的变化。例如,牛奶 α 乳清蛋白无疑起源于在哺乳动物出现之前的溶菌酶。而且已知,通常每种主要的蛋白质只有一份基因拷贝(考虑到二倍性,更确切地说,是两份相同的),但在环境压力下图景发生变化:"几乎致命"剂量的毒剂可能促使负责消除该毒剂的基因的拷贝倍增。然后这些拷贝开始发生随机突变,再选择……

蛋白质的进化得到了它们的域结构的帮助。已知,域的基因可以作为一个整体从一个蛋白质漂移到另一个蛋白质——有时与其他域的基因联合作为组合体,有时是单独地;种属接近的域常常在不同的蛋白质中、或者在独立的蛋白质中观察到[例如,钙调蛋白和小白蛋白等的钙结合域;各种三环域(kringle)等]。可能的是,基因的这类交换是由它们的内含子-外显子结构所促进的(这尤其是在免疫球蛋白中可以很清楚地看到);不过,关于交换的"模块"主要是内含子而不是整个域

①Max F. Perutz(1914—2002):英国蛋白质晶体学家,因发现血红蛋白分子结构与 John C. Kendrew 共获 1962 年诺贝尔化学奖。他也是剑桥大学分子生物学实验室的创建人与第一任主任。

的假说,看来还缺乏支持。

■ 现在来看第二个问题:我们是否看到了蛋白质空间结构的“宏观”进化,即蛋白质的结构是否随着生物体的复杂化而复杂化。看来需要回答为否定的。对蛋白质结构的观测表明,肽链的同一个折叠模式(特别是显示于图 15-1 中的)既存在于高级生物(多细胞和单细胞真核生物),也存在于细菌-原核生物中(虽然最“流行”的模式的出现频度一方面与真核生物有所不同,另一方面与原核生物也不同),即我们没有在例如细胞水平、染色质和细胞器水平,直到核糖体水平观察到蛋白质随生物体的复杂化而出现复杂化[甚至相反的是:我们看到,结构上最简单的纤维蛋白对高级生物比原核生物(尤其是太古细胞)更为典型]。

确实,存在有某些模糊的数据,表明真核细胞的蛋白质和蛋白质域比具有同样的肽链折叠模式的原核细胞蛋白质要稍微大一些,但这个问题仍然不够清楚。

不过,存在有另一种重要的“宏观”结构差别,但是这不涉及肽链折叠模式。真核生物(尤其是多细胞真核生物)的蛋白质,更倾向于共翻译修饰和翻译后修饰(糖基化、碘化作用等等)得多。修饰位点是由蛋白质一级结构所标记的,而修饰作用本身则由特定的酶所执行;而且常常只是部分执行,这导致同一蛋白质出现多种不同的形式,虽然它们的活性通常并不改变。还有,可替代剪接(alternative splicing)是真核生物所固有的,对蛋白质的多样化也有贡献。

课堂里的声音:不管怎么样,真核生物的蛋白质不仅总是尺寸较大,而且所含的域也比原核生物的蛋白质较多(典型的真核生物蛋白质包含 4~5 个域,而原核生物包含 2 个域)。

讲课人:确实如此。然而,有可能的是,关于真核生物蛋白质的尺寸较大这个一般结论,更多地与高等生物体中出现诸如免疫球蛋白这样的“外在”多域蛋白质有关,而不是与那些与细胞内部的“持家”蛋白质的变化相关。而且还需要指出,关于蛋白质链的折叠模式的“宏观进化”的问题由于基因的水平漂移的可能性而更加麻烦,因为这会导致“新的”蛋白质可能进入到“老的”生物体中。

课堂里的声音:我想回到您提到的蛋白质结构的“漂移”和它们的起源问题上来。有假说认为,蛋白质进化过程中的“跳跃元(jumping element)”可能不是整个域,而是由大约 10 个残基组成的序列。因为具有一级结构明显相似性的蛋白质的架构,其区别相当经常地只是在于肽链上额外的(或是丢弃,或是移动到蛋白质球上其他位置)小块上。而由于这种小块常常包含 α 或 β 结构,则明显同源的蛋白质域有时归属于不同的肽链折叠模式。因此,蛋白质的“漂移”包含了从一个折叠模式到另一个折叠模式的转换。要是这样的话,蛋白质不是有可能产生于这种小模块的缔合吗?

讲课人:这些问题被时不时地反复讨论。对它们还没有达成明确的回答。我们所讨论的结构模块是如此之小,而它们的序列是如此的变化多样,以至于不可能

证明它们的亲缘性。只有那些与特定功能(如与血红素相结合)有关的短序列是足够相似的。不过,功能中心的相似性有时只有当它们(如水解催化中心)在肽链折叠模式绝对不同的蛋白质中、由在蛋白质中沿肽链上相距较远的残基形成时才能观察到。在后一种情形下,很难说这些中心是从一种蛋白质"转移"到另一种蛋白质的。因此,回到局部的功能和结构模块:它们有可能是从一种蛋白质迁移到另一种蛋白质的;也有可能是在每一个蛋白质家族中重新产生的。

然而,更确切一些,应该说,这些问题还没有得到回答是对于球状蛋白而言的。在本课程中多次讲到过的纤维蛋白的序列,确实看上去像是短片段的多次重复体。因此关于纤维蛋白的"模块化"起源假说看来完全得到了证明:更何况这种蛋白质的重复结构模块常常是由单独的外显子所编码的。

然而,有理由相信,当前发展迅速的基因组计划,尤其是结构基因组计划,将会对这些令人激动的问题给出答案。

■ 我们现在已经有了大量的已整理好的蛋白质结构档案,可以提出一些哲学性的问题:①蛋白质链的典型折叠模式的简单性和规则性的物理原因是什么?②为什么同样的模式出现在完全不同的蛋白质中,而且为什么恰恰是这些模式如此引人注目?

翻译成更科学的语言,我们是想弄清楚,根据我们已经研究过的蛋白质分子物理学,肽链的哪些折叠模式是最有可能的,它们有多少,以及它们是否与在天然蛋白质中观察到的那些模式相吻合。为了寻找答案,我们首先来研究不同结构的稳定性。这样的研究途径(首先研究稳定性而不是自动折叠过程的路径)的合理性在于:同样的蛋白质空间结构可以作为在动力学上完全不同的过程的结果而获得。它们可以**在体内**(*in vivo*)[在核糖体上的蛋白质生物合成过程,在蛋白质跨膜输运(以或多或少的解折叠形式)过程中]情况下,也可以在**体外**(*in vitro*)情况下,当整个蛋白质链从解折叠状态重新折叠而获得。这意味着,行为的细节顺序在蛋白质折叠中并不起关键作用。

从简单的问题开始:为什么存在有球状蛋白的分层构造,这在我们上一讲中讨论过的。换句话说,我们来讨论,为什么紧密蛋白质球的稳定性要求蛋白质分子的骨架看上去是 α 层和 β 层的紧凑拼装,使 α 段和 β 段从球体的一个边缘伸展到另一边缘,并使不规则链段**不**处于球体内部。

大体上,我们已经讨论过这些问题了。这里关键在于氢键:由于氢键的键能很高,因此在稳定结构中应该饱和。在每一个氨基酸残基的肽基团中都有氢键的供体和受体。它们可能通过与水结合、或者形成二级结构而达到饱和。因此在稳定的(如果它们想这样的话)蛋白质球中只有二级结构有权(而非规则环线没有权)不与水接触而处于球体内部。那些含有自由的(未形成分子内部氢键)NH 和 CO 基团的元件(环线、β 片边缘和 α 螺旋末端)应该处于表面。

为了球体的稳定性起见,伸展的 α 和 β 结构应该从所有方向紧密地包围由这

些链段的侧链基团所形成的疏水核心,并且完全把它与水隔开。同时 α 螺旋和 β 片不能在混合处于同一层中,否则会破坏 β 片边缘的氢键。这意味着,蛋白质球的稳定性需要分别独立地形成 α 层和 β 层(图 15-1)。换句话说:分立的 α 层和 β 层是球状蛋白结构的稳定元件,而混合在同一层中的 α 和 β 结构会成为蛋白质球的结构性的能量缺陷。同时很明显,**稳定的**蛋白质球应该,至少主要地,由稳定元件所组成,而且它应该避免任何能量缺陷。而由于我们只能观察到稳定的蛋白质球(不稳定的无法观察到,它们解体了……),则在蛋白质中被观察到的结构元件应该主要是稳定的,缺陷应该是罕见的。

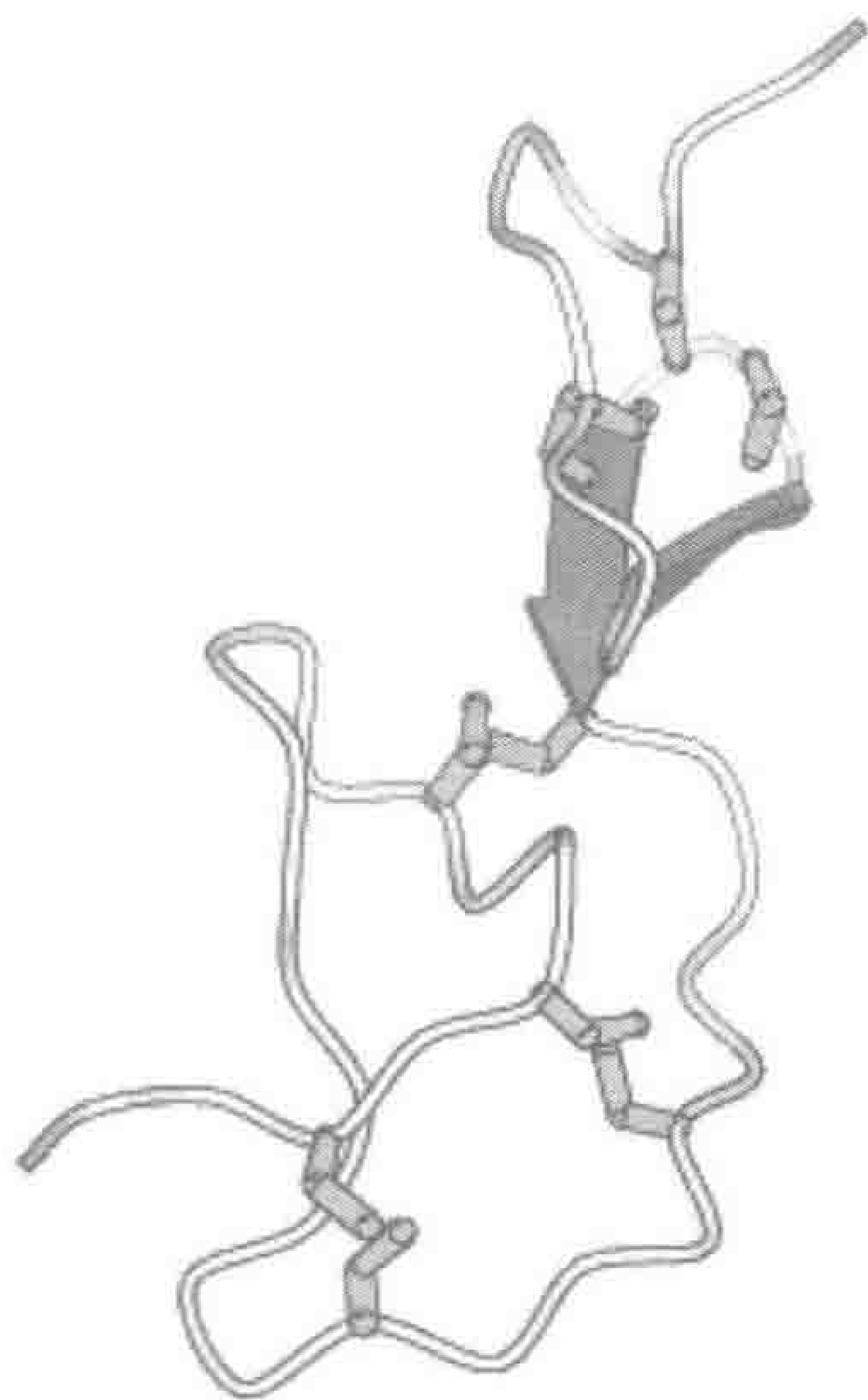


图 15-3 罕见的既没有 α 结构、也几乎没有 β 结构的蛋白质球(huristasin 蛋白,"贫二级结构"类蛋白质的代表)。这个蛋白质的序列特殊,含有大量的能形成二硫键的半胱氨酸(Cys)(它们的侧链基团显示为灰色的弯折粗棍)

特别是对 α 和 β 层来说,确实如此。它们是稳定的。而正如我们已经看到的,这些层(通常不是平坦的,而是扭曲的,有时候是圆筒形的,而在 α 螺旋球体中,甚至是准球状的),而且是"未混合的"层,对蛋白质球是很典型的。

分层结构简化了蛋白质的构造。

绝大部分域可能呈现为二层、三层甚至四层(少见)结构。

单独的蛋白质(尤其是那些含有金属有机复合体或大量二硫键的侧链基团的)可能并不完全满足这种"分层拼装"示意图(图 15-3),但这样的蛋白质非常少。

多于四层的域实际上不存在,而且在原理上也很清楚是什么原因。在它们的与水隔离的内部需要放置太多的残基,而蛋白质链(更准确些,是水溶性球状蛋白)中非极性和极性残基的典型比率是 1 :1,大量的极性残基不得不居留在蛋白质内部,这是能量极其不利的:这样的蛋白质不可能是稳定的。因此由"普通"氨基酸组成的极大的(从而,很多层的)单一蛋白质球应该是不稳定的,所以大蛋白质需要分割为多个亚球,即多个域。

确实,几乎连续地由疏水氨基酸构成的肽链可以形成极大的稳定蛋白质球。但是,首先,这样的序列比混合的"疏水-亲水"序列要少得多,其次,这种序列喜欢进入到细胞膜中,而不是形成水溶性球状蛋白。

■ 在原则上,似乎可以虚构一种氨基酸序列,其侧链基团能"治好"所有"缺陷"和所有主链和水之间氢键的断裂(这些断裂要求 β 片或环线的边缘浸入到球体中),并且以氢键喂饱喜欢处于蛋白质内部的极性侧链基团。或者,虚构一种序列,能以强有力的连接[例如,二硫键(S—S)或配体键(就像在金属有机化合物中)]补

偿那些氢键断裂还绰绰有余。可以这样虚构。但这将是**非常特别的**、意味着**非常罕见的**序列……

也许,这就是问题的实质,也许,“普通的”球状蛋白是由“普通的”(即相对不那么严格地挑选出来的)序列构成的,而不是由“非常罕见的”(即非常严格地挑选出来的)序列构成的!

■ 让我们来看一看蛋白质的一级结构(图 15-4)。统计分析表明,水溶性球状蛋白(我们现在所讨论的)的氨基酸序列看上去像是“随机的”,即在这些序列中各种氨基酸残基大致就像是在随机共聚合反应(co-polymerization)中所能期待的那样被混合起来。当然,每一种序列都不是随机生物合成的结果,而是由基因编码的。然而,水溶性球状蛋白的氨基酸序列看上去像是“随机序列”。这里的意思是说,在它们中既没有膜蛋白所特有的(其中有很明显的必定处于膜中的疏水块与必定形成由膜上伸出的环线和完整域的亲水块的交替)**模块性**,也没有纤维蛋白所特有的**周期性**(它们具有巨大的规则二级结构)。

蛋白质	序列	
球状蛋白	oo•oo••oo•o••o•oo•ooo••o•oo•o•oo••o•oooooo•oooo•	准随机的
膜蛋白	•••••••••••oo•oooo•ooooooo•••••o••••••oooo••oooo•o•o | 疏水性 ‖ 亲水性 |	模块性的
纤维蛋白	•oo•ooo•oo•ooo•oo•ooo•oo•ooo•oo•ooo•oo•ooo•oo• | 周期 |	周期性的

图 15-4 在水溶性球状蛋白、膜蛋白和纤维蛋白的一级结构中的疏水氨基酸(•)和极性氨基酸(o)交替的典型模式

课堂里的声音:不能不注意到,编码后的信息也同样可能看上去像是随机字母序列,虽然绝非如此……

讲课人:当然,球状蛋白的氨基酸序列并不是真正随机性的(意思是,否则所有的序列都能折叠成球状蛋白)。当然,这些序列被挑选出来是为了形成稳定的蛋白质球。但是蛋白质球的形状变化多样。因此在所有一级结构组成的集合中,包含了所有这些蛋白质球所固有的规则性的完整谱(即各种不同密码的巨大集合)。在我们说这些一级结构是“随机的”(更确切地说,是“准随机的”)的时候,其意思仅仅是指,在所有球状蛋白的一级结构的总和中,挑选能形成蛋白质的氨基酸序列的选择痕迹不能清楚地跟踪(因此不能那么强烈地限定可能序列的大量性),就像在纤维蛋白中以周期性为选择痕迹,或在膜蛋白中以模块性为选择痕迹。正是在这个意义上我们说,水溶性球状蛋白的氨基酸序列看上去大致就像随机序列。

而“看上去像是随机序列”是什么意思?这意味着,看上去像是所有可能序列中的**大部分**……意味着,在讨论水溶性球状蛋白时,可以完全有理性地提出这样一

个问题:哪些稳定的空间结构通常是由最普通的、随机的或者与之类似的("准随机的")序列所编码的?

还有。如果在蛋白质球中含有"结构缺陷"(如不规则的环线或 β 片边缘浸入在疏水核中),则为了保障这种球体的稳定性,需要特别苛刻地挑选氨基酸序列(以便在其中有更多有利于稳定性的相互作用)。而且蛋白质结构的"能量缺陷"越大,挑选应该越苛刻。而如果没有缺陷,则不需要苛刻的挑选,即"无缺陷的"蛋白质球能由相当多的序列所稳定,缺陷不大的球体(序列不多,缺陷很大)序列数量非常少。

而且,如果只考虑物理学,则被大量序列所编码的结构应该能经常观察到,而由少量序列所编码的只能很少见到。蛋白质结构的"物理学挑选"可能就是这样出现的。

由于我们已经弄清楚,在球状蛋白中的二级结构的典型拼装和"堆码"(图 15-1)看来正是随机或几乎随机的氨基酸序列的稳定拼装包所应该看上去的样子,则我们更加确信,至少在**拼装**水平,所观察到的它们的自然(生物)选择与物理学选择并不矛盾。

■ 让我们更进一步来讨论蛋白质链的**折叠模式**:以上述同样的观点,从主要是"无缺陷的"(因此对大量序列都"适宜")结构的能量缺陷和物理学选择的观点来讨论。

正如我们已经看到的,蛋白质链的拼装模式常常令人惊讶地美丽。蛋白质链的通路常常好像装饰陶器的线条(图 15-5)。而且,根据发现这种相似性的 Jane Richardson① 的深刻思想,这不是偶然的,装饰线条和蛋白质链"解决"的是同样的任务:在避免线条自身相交的条件下包围某个体积(在蛋白质中就是球中心,它的疏水核)。

在蛋白质中,这种效果是这样实现的:结构段围绕核(或两个核,后者在 α/β 型蛋白质中是典型的)而排布,而环线沿核表面滑行:通常,它们连接彼此**反**平行(而**不是**平行)相邻的二级结构段(图 15-6a),而且**互不交叠**(图 15-6b)。注意到,后者实际上消除了在蛋白质链中出现打结。

为什么以环线相连的平行连接比反平行连接要差?有可能是因为这需要太长而且不规则(即没有以氢键支撑)的环线?或者有可能是刚性的多肽链必须要更急剧地弯曲并在它上面花费能量?

而环线的交叠又有什么不好,要知道在这种情形下肽链不是互相嵌入,而仅仅只是其中一个在另一个上面跨过?有可能,"下面的"环线紧贴着核,失去了与水的氢键?为了补偿这种损失(这种"能量性缺陷"),再一次需要"罕见的"序列……

---

①Jane S. Richardson (1941-),美国 Duke 大学生物化学教授,她的一个重要贡献是发明了用于表现蛋白质三维结构的条带(ribbon)图示法。

应当指出，在自然界中观察到了带有如环线交叠这样相对较小缺陷的蛋白质结构（这与带有很大缺陷的结构，如 α 层和 β 层混合，不同）。但它们非常罕见，而这种缺陷的罕见性尤其具有重要意义，因为把肽链排布成“有缺陷的结构”可能比排布成“无缺陷的”结构的方法数量要多得多。

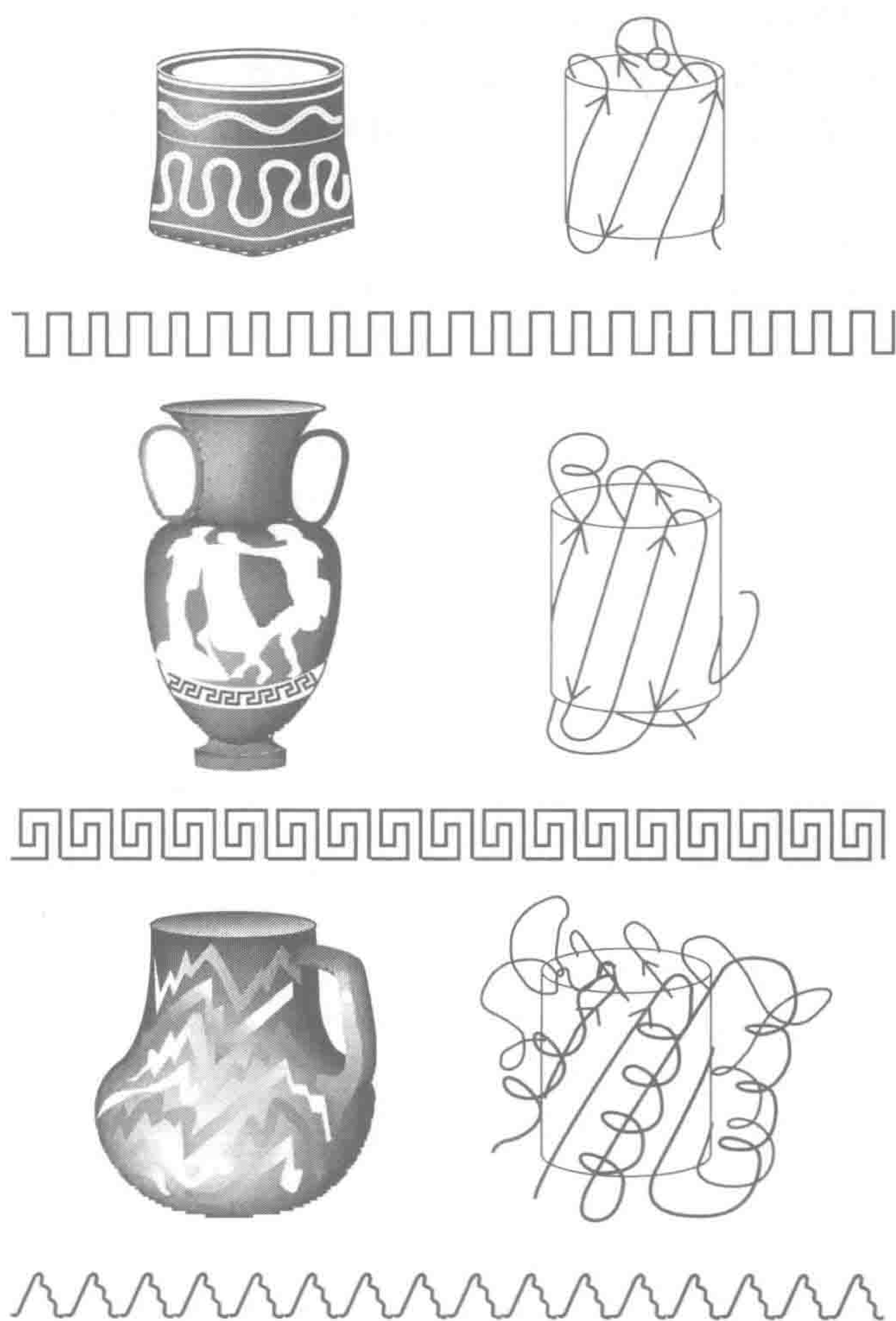


图 15-5 蛋白质链的折叠模式与在印第安和希腊花瓶上的装饰线：用自身不相交的线条包围体积的两种解决方法。上：河曲模式；中：希腊钥匙模式；下：“闪电”锯齿形模式。图片取自 Nature，v. 268，No. 5620，1977（© 1977，Macmillan Magazines Limited）的封面（经允许），该期杂志刊登了 J. Richardson 关于蛋白质链折叠模式的文章

因此，“结构性缺陷”强烈地影响不同蛋白质结构在自然界中的出现频度。在这里，我们确实会感到困惑：在环线交叠型的“缺陷”中，只是减少 1 个，最多 2 个氢键，即能量损失并不大，3kcal/mol 或 5kcal/mol。这不仅比蛋白质中的相互作用的

总能量[通常以几百千卡计算(根据蛋白质熔化实验得出)]小得多,而且很明显小于通常的蛋白质“稳定性储备量”(即它的天然态与变性态的能量差值)。在天然条件下这种储备量约为(根据同样的实验)10kcal/mol。为什么只值 5kcal/mol 的“缺陷”会禁止,或者几乎禁止,天然蛋白质球中的环线交叠?

还有一个问题:是什么妨碍了在环线中进行额外的弯曲(图 15-6 中的虚线)以避免环线与环线的物理交叠(即以旁通方式取代环线交叠)?也许,这里的关键又一次在于多肽链的弹性,因为在额外弯曲上环线需要花费(正如估算所得出的)几个(又是几个!)千卡?

这些问题稍迟一点再回答,而现在来讨论蛋白质架构的另一个典型性质:在平行 β 串之间的连接线几乎总是与它们形成**右**旋螺旋,而不是**左**旋螺旋(图 15-7)。

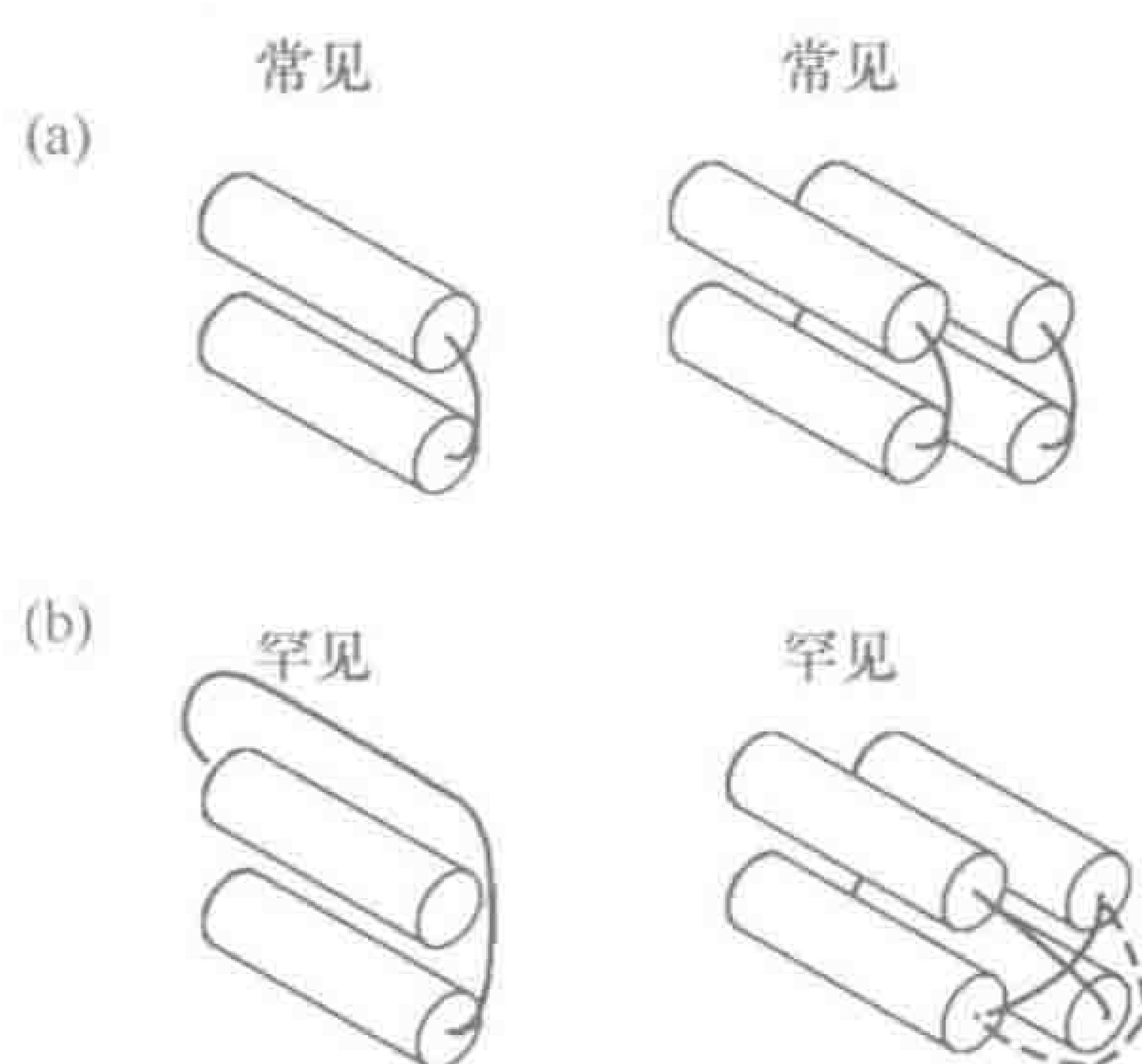


图 15-6 通常,环线连接彼此**反**平行(而**不是**平行)相邻的二级结构段(a),而且在蛋白质中极少观察到环线的交叠,无论是一条环线从另一条的侧旁通过,或者一个环在另一个环之上(b)

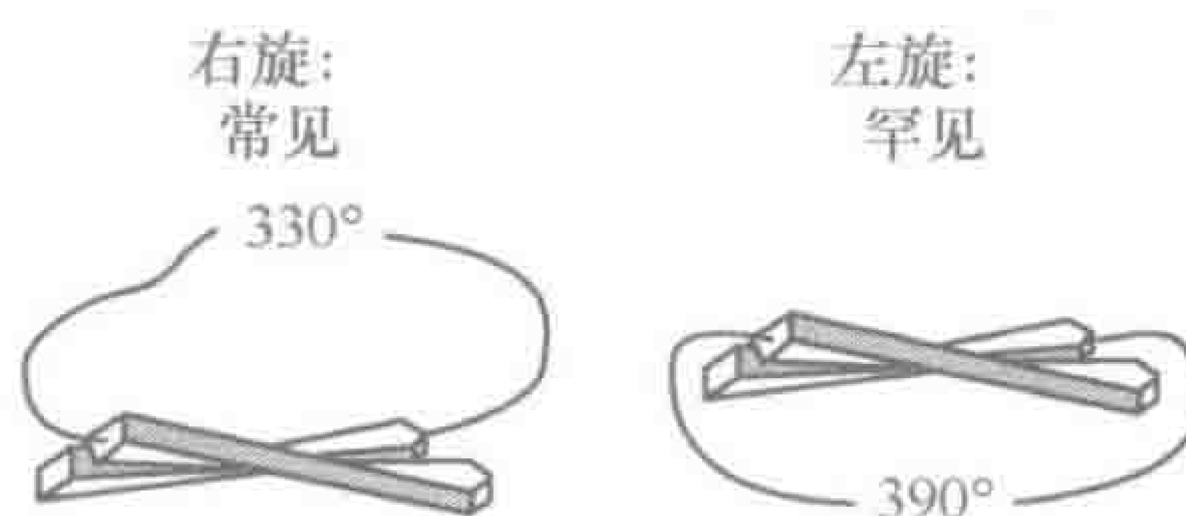


图 15-7 平行 β 串之间的连接线的左旋通路在蛋白质中极少观察到,而右旋的则很常见(连接线在图中简单地绘制为线条,实际上通常还包含 α 或 β 结构段)。在右旋通路中肽链以逆时针方向运动而接近我们(沿螺旋圈的轴观察),在左旋通路中按顺时针方向

在这种情况下,稳定性判据让我们能够在两个镜像对称的蛋白质结构中选出“更好的”。差别是建立在天然氨基酸的镜像不对称性的基础之上的。正如你们记得的,它导致了由 L 型氨基酸组成的 β 层以右旋扭转占优势(图 15-7 所示)。这时相邻 β 段的轴之间的角度约为 30°,从而使右旋连接线的总转动角约为 330°,左旋的约为 390°。结果是,由于肽链的刚性,右旋连接线更有利,它的自由能较低,但又一次并不大,只有几个 kcal/mol。

在这里,任何一个懂得高分子聚合物物理学的人都应该打断我并且说:“高分子聚合物的弹性是熵效应,而不是能量效应!即强弯曲的链不可能像直链或弱弯曲的链那样自由地变动。换句话说,具有强弯曲形状的链的构象比具有比较伸直形状的链少得多。然而你讲到的效应涉及的是变动的链,即**没有**固定在球体中的链。但是在天然蛋白质中,肽链是固定的,而且无论它如何走向,它总是只具有某一种构象。你所说的熵损失与蛋白质的天然结构(其中的肽链总是固定的,即总是熵为零)之间有什么关系呢?”

回答这个问题(关于链过于急剧弯折的熵性缺陷)并不复杂。在此我们来给它一个定性的回答。这里的问题在于,肽链构象的“受限性”使之无法在组成它的氨基酸序列之上建立起来。这样的构象将只对于严格确定的序列(或狭窄的序列集合)具有低能量,而对其他序列,即大部分序列,具有高能量。因此熵效应转化为能量效应。而如果肽链可以从大量构象中挑选,则几乎每一个序列都能从中找到自己可以接受的某一种构象。

在下一讲中我们将更严格地讨论关于能量缺陷的问题。我们还将讨论遗留下来的另一个问题,为什么能量值总是只有几个 kcal/mol 的“缺陷”在蛋白质结构的出现频度上起着如此本质性的作用。

而现在,还不能回答最后一个问题,但是在定性的意味上,让我们讨论一下在蛋白质结构上提到的其他统计学规律与能量如何相关。例如,疏水和亲水侧链基团浸入到蛋白质球中,肽链中不同的转动角,如此等等。结果是,与蛋白质链的折叠模式所讲到的完全相同:“缺陷”出现得很少,而“好元件”经常出现!但这里我们列出更多的统计学公式,并有可能得到不仅是定性的,而且是定量的估计。

所有这些估算得出如下结果。“缺陷”(无论是“差的”内部转动角或极性基团深埋到蛋白质球内)出现频度的降低,和“好元件”(如带异种电荷的基团形成的盐桥)出现频度的升高,都可以用普适的唯象公式描述:

$$\text{出现频度} \sim \exp[-\text{元件的能量}/(kT_C)] \tag{15.1}$$

式中,T_C 为接近于室温或蛋白质典型熔化温度的温度:因为“在数量级上”300K 与 370K 是相同的。

用于描述蛋白质中的结构细节出现频度的式(15.1)与玻尔兹曼统计公式惊人地相似。确实,在**指数形式**方面很相像,但在物理意义上不相像。需要提醒的是,玻尔兹曼统计公式的基础在于,粒子在各个位置上游荡,每一个粒子在其能量较低的位置停留的时候更长。然而在所观察的(天然的)蛋白质球中,所有结构元件都是固定的,它们**不能**出现和消失,**不能**从一个地方跑到另一个地方!

也就是说,通常的玻尔兹曼统计并不能在这里应用于蛋白质细节结构的出现频度上,完全不行。既然是这样,那为什么蛋白质中的这些细节的出现频率统计公式具有这样熟悉的“准玻尔兹曼”形式?

我们把对这个问题的详细解答留到下一讲,而现在把它作为一个现象性的事

实来接受;能量值为 kT_C(即大约为 1kcal/mol)的缺陷,会把含此缺陷的结构的出现频度降低到几分之一。而且,正如我们已懂得的,这意味着,这种缺陷会把能使含此缺陷的蛋白质维持稳定的氨基酸序列的数目降低到几分之一。

■ 于是,我们得到了如下结论。

(1) 蛋白质链的"流行的"折叠模式看上去是如此"标准",即如此简单和规则,是因为蛋白质结构的框架是伸展的标准固体(α 螺旋或 β 串)的紧凑分层拼装,而不规则的连接线沿球体表面走过,既不彼此交叠,也不与结构片段的末端交叠。这种排列的物理学原理在于它对蛋白质球的稳定性是最为有利的,因为它保证了非极性侧链基团与水隔离,而且所有主链肽基团甚至在包埋到密实蛋白质球中时都饱含氢键。

(2) 肽链的这些"标准的"稳定折叠模式的数目相对并不多(约为几百个,而蛋白质数以万计);因此,并不奇怪的是,一些这样的"通用"结构出现在从各方面来看都不相同的蛋白质中。

(3) 同时,其他的"有缺陷的"蛋白质链折叠模式并不是禁止的,它们只是可能性很小(即只有相对很少数目的序列能保证它们的稳定性),而且缺陷越大,则可能性越小。

(4) 看来,对于具有"准随机性"氨基酸序列的水溶性球状蛋白的域的结构,可以提出"多数原则":**能够满足给定的架构而不破坏其稳定性的氨基酸序列越多,则这种架构在自然界中出现得越多**。

第 16 讲

对随机和准随机氨基酸序列，可以期望有哪种二级结构？准随机长序列的稳定空间结构的域构造最有可能。蛋白质结构细节的准玻尔兹曼统计。准玻尔兹曼统计来自于对稳定的蛋白质结构的物理性选择。结构元件的稳定性对球状蛋白未破坏的空间结构的一级结构的选择严格性的影响，或者说：为什么一种蛋白质结构经常出现，而其他的很罕见，在大蛋白质球中心哪一种结构，α 或 β，更值得期待？"熵性缺陷"与"能量缺陷"之间的联系。球状蛋白是作为"中选"的随机肽链而出现的吗？蛋白质工程中"蛋白质有利的"随机序列的选择。

现在我们来更加详细地讨论在蛋白质结构中观察到的基本规律与这些结构的稳定性和编码它们的氨基酸序列数量之间的关系。

请看图 16-1，回忆一下，典型的蛋白质球体的框架为伸展的规则固体（α 螺旋和 β 段）的紧凑分层拼装，而不规则连接线沿球体表面走向，既不相互交叠，也不与结构片段的末端相交。

前面讲到，蛋白质这种构造的物理原理在于，它允许非极性蛋白质基团与水隔离，而主链的肽基团甚至在包埋到紧凑蛋白质球中时也能使自己的氢键饱和，从而对球体的稳定性最为有利。而这种"对稳定性有利"也允许更多数量的不同氨基酸序列能适应于给定的球状架构，而不破坏其稳定性（这就使得这种架构在自然界中更频繁地被观察到，我们称之为"多数原则"）。

正如我们已经讲到过的，典型的水溶性球状蛋白的"独特之处"在于，在它们的一级结构中不存在任何显著的共同性质（而且因此这些序列最大限度地各不相同、数量众多，这又使得可以以"多数原则"的观点来讨论它们）。确实，在它们的一级结构中找不到任何明显的相关性：没有纤维蛋白那样特有的周期性，也没有膜蛋白那样特有的块交替。相反，在水溶性球状蛋白中，极性基团与非极性基团相当均衡地"混合"。而且这些蛋白质中的极性和非极性残基的数量也大致相等。结果就是，水溶性球状蛋白的一级结构完全像是由疏水氨基酸和亲水氨基酸随意合成的"随机共聚物"。

■ 但是，"随机"序列与肽链紧凑折叠成球相容吗？它们与在蛋白质中观察到的二级结构（通常占据了蛋白质链的半数以上）相容吗？要回答这些问题，让我们来看假想的由相同数目的极性基团和非极性基团的随机共聚反应得到的"蛋白质链"。

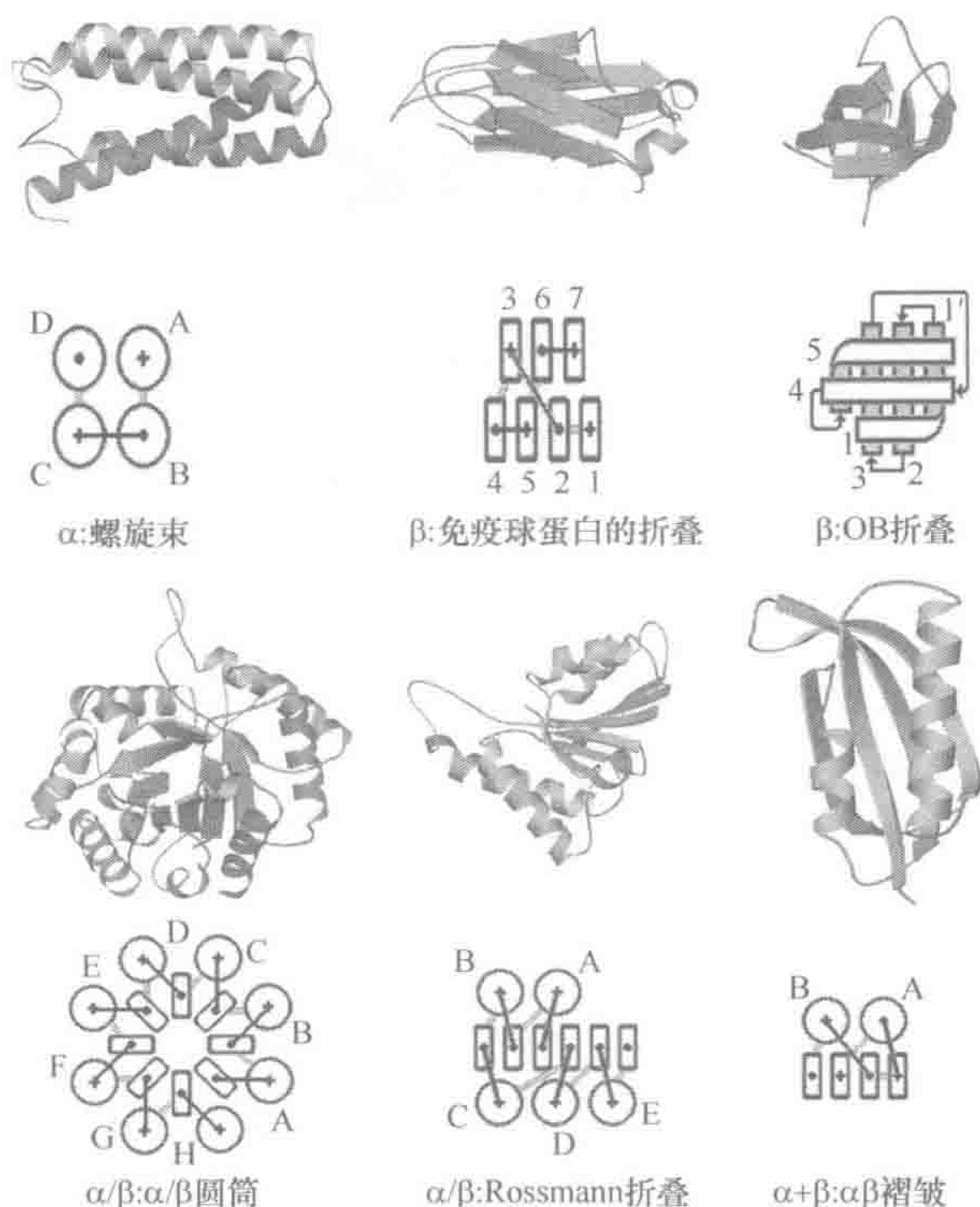


图 16-1　在 α,β,α/β 和 α+β 型蛋白质中的典型蛋白质链折叠模式。下方为它们的简化示意图。注意 α 结构和 β 结构的分层拼装,以及每一层要么只由 α 螺旋组成,要么只由 β 串组成,但不是同时由 α 螺旋和 β 串组成

一个 α 或 β 结构片段要能够被装进紧凑球体,它应该具有连续的疏水表面。在 α 螺旋中,这样的表面由占据相对位置为 i - i+4 或 i - i+3 的非极性侧链基团所形成,而在 β 串中为 i - i+2。容易证明,甚至在随机共聚物中都具有相当周期性的非极性基团聚合体(图 16-2),能够形成可以装进中等大小的蛋白质域典型球体的 α 或 β 片段的疏水表面。

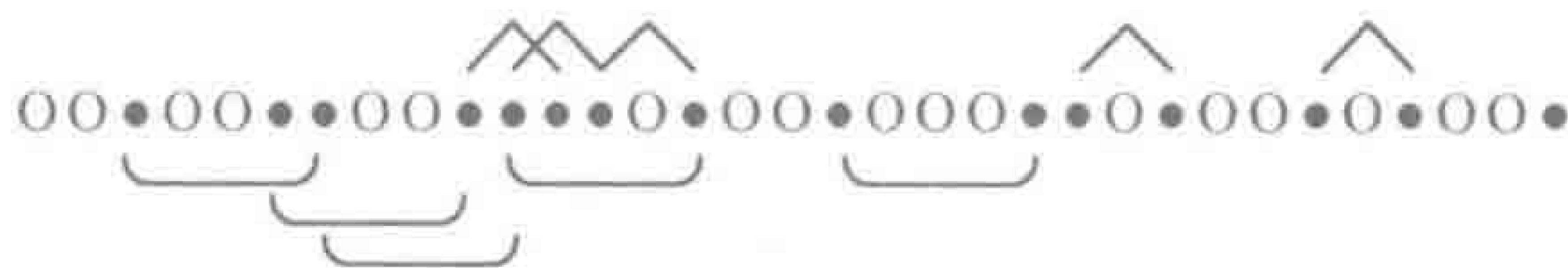


图 16-2　在水溶性球状蛋白的"准随机"一级结构中的疏水氨基酸(•)和极性氨基酸(o)典型交替模式。弧线和折线标出了在该一级结构中存在的适合于 α 螺旋段和 β 结构段的疏水表面的排列位置,即分别为 i - i+4 和 i - i+2 的成对疏水残基

假设非极性基团在共聚物中的比例为 p,极性基团的比例为 $1-p$。则在链上给定点处开始、恰好由 r 个非极性基团组成(就是说,限制为开头和结尾是两个极性基团)的周期性序列的概率为

$$W(r) = (1-p)p^r(1-p) \quad (16.1)$$

关于 α 片段和 β 片段的“疏水表面”，暗含的意思是它包含超过一圈的相应结构。或者说，在 α 螺旋的疏水表面中在周期性位置 $i - i+4-\cdots$ 中分布着 $r\geqslant 2$ 个疏水基团（目前我们不把注意力投向在它们之间的残基，因为它们并未参与到所讨论的螺旋表面）。在 β 串的疏水表面中在周期性位置 $i - i+2 -\cdots$ 中分布着 $r\geqslant 2$ 个疏水基团（图 16-2）。把式(16.2)应用于这些周期性位置，我们会看到，疏水表面所覆盖的 α 片段和 β 片段的平均期望圈数为

$$\langle r\rangle = \frac{\sum_{r\geqslant 2} W(r)r}{\sum_{r\geqslant 2} W(r)} = \frac{\sum_{r\geqslant 2} rp^r}{\sum_{r\geqslant 2} p^r} = 2 + p(1-p) \quad (16.2)$$

（我擅自省略了对级数 $\sum_{r\geqslant 2} rp^r$ 和 $\sum_{r\geqslant 2} p^r$ 的求和，所有这些都能在数学参考书上找到）。所以，“平均的”α 螺旋和“平均的”β 段都包含（在 $p=1/2$ 时）$\langle r\rangle=3$ 个规则分布的疏水基团，即 3±0.5 个完整的 α 或 β 结构的全周期（±0.5，因为疏水基团可能位于一圈的开头，或中间，或结尾）。在 α 段和 β 段（它们的周期数为 3.6 和 2）中的平均期望残基数分别为 $\langle n_\alpha\rangle=11\pm2$ 和 $\langle n_\beta\rangle=6\pm1$。这实际上与在球状蛋白中的 α 段和 β 段的平均长度相吻合。有趣的是，在随机序列中及实际蛋白质的一级结构中，能形成 α 段和 β 段疏水表面的非极性基团富集部分常常在相同的肽链段上重叠（图 16-2）。

类似的估算指出，随机共聚物中二级结构之间的环线的平均长度为 $3+0.5p^{-2}$ 数量级，即在 $p\sim1/2$ 时环线平均要比二级结构中的链段要短，这也被实际观察到了。

这样，在随机共聚物中，在 $p\sim1/2$ 时，能够自己形成连续的 α 段和 β 段疏水表面，这些 α 段和 β 段能至少有一面附着在疏水核上，而环线相对比较短。所以“平凡的”随机序列完全能够形成由至少两个二级结构层构成的折叠。

课堂里的声音：可是，不应该忘记，在某些（即使是少量）球状蛋白中（如在血凝素或亮氨酸拉链中）出现有极长的螺旋，与上述原则不符。而在另外某些（如超氧化物歧化酶）中，有极长的不规则环线……

讲课人：确实如此。在这些情形下，它们要么被看做是从螺旋纤维蛋白中借来的模块，要么（就长环线而论）被看做是异常的亲水模块。但是在**平均**上，球状蛋白中**多数**的 α 段、β 段和不规则段并不大，其长度恰好与疏水和亲水基团含量相等的“随机”氨基酸序列相同。

■ 当肽链长度不超过 150 个残基时，“随机”序列与球体的紧密的、可能最稳定的形式是协调一致的。然而，随着蛋白质球的尺寸增大，随着其中的二级结构层数的增加，它对随机氨基酸序列的“适用性”会降低。因为分布在蛋白质内部的链段应该较大（与球体直径相当），而且应该几乎完全仅由疏水残基所组成，如果存在有

大量与水隔离的亲水基团，球体就会爆裂。在由约 50% 的疏水残基和约 50% 的亲水残基构成的随机共聚物中也可能“随机地”形成少量这样的较长且几乎完全疏水的一级结构链段，但仅仅很少量。因此对于随机序列我们只能预期存在有两三层、偶尔是四层的结构片段，不能更多，这也正是在不很大的单域水溶性球状蛋白和更巨大的水溶性球状蛋白的域中所观察到的（图 16-1）。而巨大的蛋白质应该由亚球，即域，所组成。这也是实际观察到的。

■　现在我们面临的是回答在上一讲中提出的两个问题（迄今还没有给出应有的回答）。

（1）为什么只值几个 kcal/mol 的“能量缺陷”（在蛋白质总能量比这大得多的背景上）实际上可以使许多蛋白质架构的模式被禁止？

（2）“熵”效应与蛋白质天然结构（其中的肽链总是被固定的）有什么样的关系？

让我们从第一个问题：关于“缺陷”能量在蛋白质架构的统计公式中的表现形式开始。但是首先我们来加强这个问题：请回想我们已经在上一讲中说到，在定性水平上，“结构缺陷”的罕见性得到了已观察到的蛋白质结构中的“微小元件”的“准玻尔兹曼”（即“类似于玻尔兹曼的”）统计分布的支持，这也是我们需要弄懂的。所以就从它（“微小元件”的统计分布）开始。

■　作为例子，让我们来看氨基酸残基在蛋白质球内部和表面的统计分布，并看一看它是如何与氨基酸残基的疏水性相关联的。

氨基酸残基的疏水性通常以它们从模拟蛋白质疏水核的辛醇转移到水中的转移自由能来衡量。图 16-3 中纵坐标为这种转移自由能（除以 RT，其中 $T=300\text{K}$），横坐标是在蛋白质表面和内部的残基数量比的对数。可见，这些点大致排列在一条直线上，且该直线的斜率大致接近于 1～1.5。

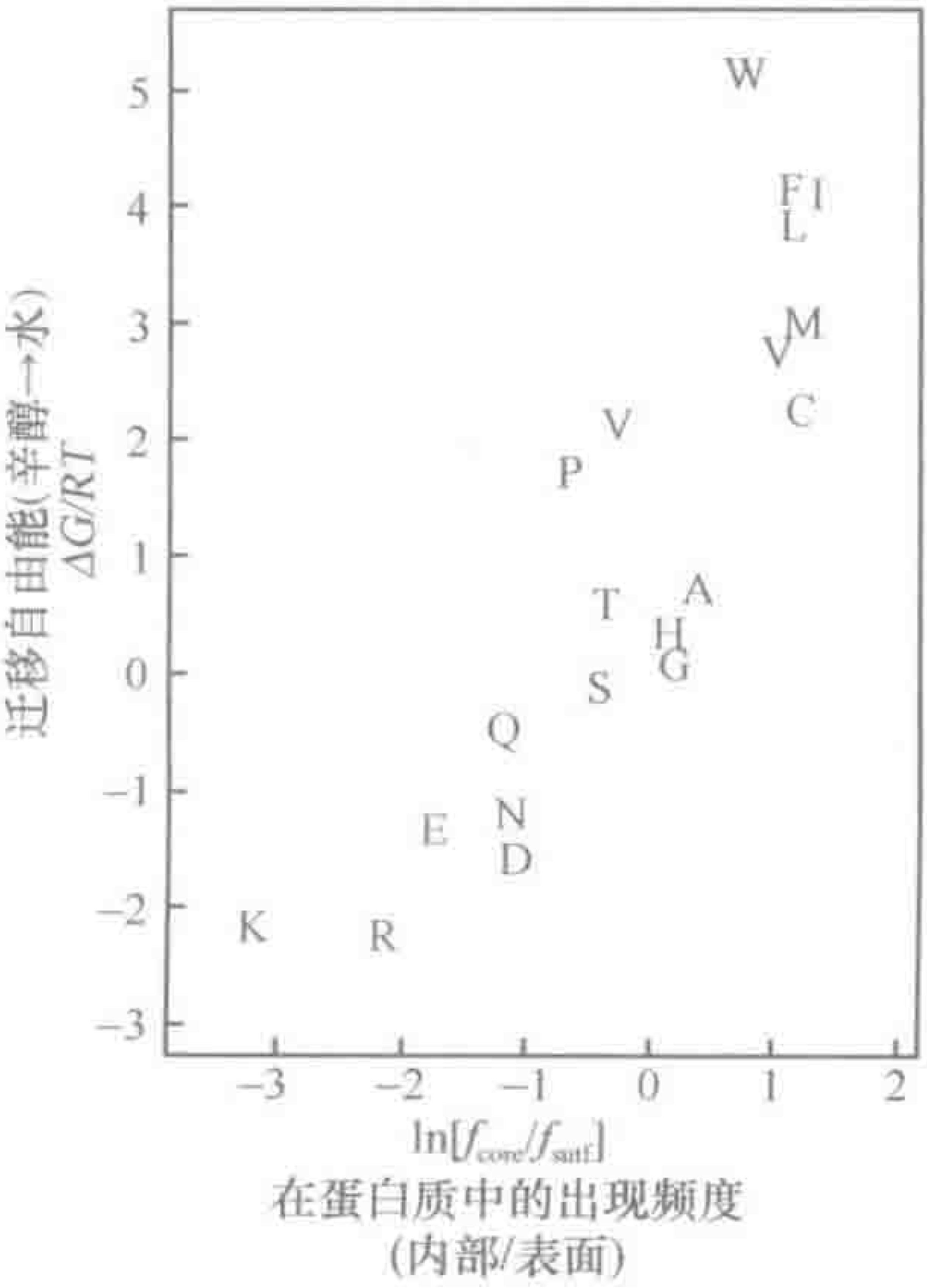


图 16-3　氨基酸残基从非极性溶剂转移到水中的侧链基团转移自由能实验测定值（ΔG，以 RT 为单位表示）与“残基从蛋白质核迁移到表面的表观自由能”（也以 RT 单位表示）。后者是从氨基酸残基在蛋白质内部和表面的观测出现频度（f_{core} 和 f_{surf}）根据公式 $\Delta G_1/RT=-\ln[f_{surf}/f_{core}]$ 计算而得。图片取自 Miller S.，Janin J.，Lesk A. M.，et al. 1987. J Mol Biol. 196：641-656，略有修改

因此，所观察到的残基在蛋白质球内部和表面的统计分布可以很不错的用下式描述：

$$出现频度 \sim \exp(-在给定环境中的自由能 / kT_C) \tag{16.3}$$

式中，"构象温度"T_C 在数量级上为 300~400K。

课堂里的声音：我建议更谨慎地对待显示于图 16-3 中的数据…… 要知道，在实验上疏水性是根据氨基酸从水迁移到高分子的辛醇而确定的。但是谁能说，蛋白质疏水核恰好正像辛醇一样？更不用说，辛醇是液态的，而蛋白质核心是固态的；难道不可以这样说。例如，纯疏水的环已烷是比辛醇更好的疏水核模型？而由于极性基团在环已烷中的溶解度非常低，必然会强烈地拉长图 16-3 中的疏水性坐标轴…… 想必，疏水性与出现频度的相对关系在这种情形下也会保持，但穿过"环已烷"的点的拟合直线的斜率可能会更大，为 3~4。还有，顺便说到，在图 16-3 的上方部分观察到也正是这样的斜率，对于分布在那里的疏水氨基酸，把辛醇更换为苯对它们的影响很小……

讲课人：嗯，在实验上测定疏水性时用什么来模拟蛋白质的疏水核心确实是个问题…… 我想说的是，辛醇终归还是要比环已烷要好；因为在蛋白质核心中存在有极性原子，我指的是围绕核心的主链 NH 和 CO 基团（确实，它们通常在围绕核心的二级结构中形成氢键，但 CO 基团能够形成两个"叉状"氢键，其中只有一个在二级结构中被使用，而另一个仍是不饱和的）。因此在纯非极性环已烷下的疏水性刻度未必需要调整。另一方面，如果说到数值估算，不应该忘记，制作图 16-3 时是把所有侧链基团根据它们在蛋白质中的包埋程度而粗略地划分为两类（"核心的"和"表面的"）；而在这种分类下"核心的"和"表面的"基团的暴露程度区别不是基团的整个表面，而大约只是它的一半。因此实验上的疏水性应当相应地**减小**一半，而这（与使用环已烷代替辛醇相反）会减小拟合直线的斜率…… 总之，我完全同意，对这里所有的**数值**估算应该持合理的怀疑，但应该把注意力集中在不同元件的能量与它们在蛋白质中的出现频度之间的定性关系上。

■ 可见，氨基酸残基在蛋白质内部和表面的出现频度的统计分布公式确实像玻尔兹曼分布——**在形式上**相似！这最初是 Pohl 于 1971 年在蛋白质中的氨基酸残基的侧链内部旋转角分布中观察到的。后来证明，蛋白质结构的许多其他元件的统计分布中也是如此：离子对的出现频度，残基在二级结构中的出现频度，孔洞在蛋白质中的出现频度等。迄今这种相似性已经变得如此司空见惯，以致蛋白质结构的统计分布公式常常被用来估算氨基酸残基之间各种相互作用的自由能。

然而，这里需要强调的是，蛋白质的统计分布公式与玻尔兹曼分布的相似性**仅仅是其指数形式**，而**不**在于物理意义。你们可能还记得，玻尔兹曼统计分布的基础在于，大量粒子能从一个地方游荡到其他地方，其中的每一个粒子都在其能量较低的地方停留更长时间。而在天然蛋白质中氨基酸残基**不是**从一个地方游

荡到其他地方！例如，抹香鲸肌红蛋白肽链中的 Leu72 总是处于这种蛋白质天然结构球体的内部，我们永远都不会在表面看到它。尽管我们看到，根据统计学，所有亮氨酸（Leu）中的 80%~85% 处于蛋白质内部，15%~20% 处于表面，但这并不是说每一个亮氨酸有 80%~85% 的时间处于球体内部而 15%~20% 的时间处于表面。这里指的是，自然选择把绝大部分亮氨酸固定在那些处于蛋白质球体内部的肽链位置上。

这就是说，普通的玻尔兹曼统计式是在普通三维空间中的涨落统计，与残基在蛋白质核心和表面的分布毫无联系。在球体中并没有涨落能把每一个亮氨酸（根据其疏水性）都搬到蛋白质表面呆上 15%~20% 的时间，又拿到蛋白质内部呆上 80%~85% 的时间，即对于每一个单独的亮氨酸不存在由其疏水性所确定的玻尔兹曼分布。既然如此，那么如何解释所有全部亮氨酸在蛋白质内部和表面的出现频度与由其疏水性所确定的玻尔兹曼分布相一致呢？

从另一角度来看。例如，亮氨酸主要分布在蛋白质球体内部有什么好处？因为这提高了它的稳定性。但是那么为什么自然选择并没有把所有亮氨酸都固定在蛋白质内部？也许，是因为 80%~85% 的亮氨酸处于内部已经**足够**保证蛋白质的稳定性了，而处理其余那些需要太多的作用力了？

让我们把自然选择的心理学作为不得要领和非学术的问题放在一边，来看另一个问题：蛋白质结构中某种元件的内部自由能如何改变有助于加强含有该元件的蛋白质的稳定性的**氨基酸序列的数量**。

■ 例如，让我们来看，在蛋白质内部某点上发生亮氨酸（Leu）→丝氨酸（Ser）的突变时，蛋白质的稳定结构的数量变化如何。

如果蛋白质的天然结构的自由能既比解折叠的结构要低，也比该链的其他任何球状折叠要低，即如果天然结构的自由能处在比任何与天然结构无关的界限之下，则天然结构是稳定的。为了简便起见，我们认为，天然结构只与链的解折叠状态竞争，而且①残基对蛋白质稳定性的贡献取决于它的疏水自由能（这并不完全准确，至少还有氢键和必须紧凑拼装，但以“零”近似来忽略它们……）；②蛋白质内部残基完全与水隔离，而表面的残基完全与水接触（这是相当粗糙的，但是在定性上是可靠的假设），而且变性蛋白质是完全解折叠的（常常是这样，但并非总是如此；因此下面所述的理论是近似的，然而却是简单的）。

Ser 的侧链基团从疏水环境到水中的转移自由能大约为 0，而 Leu 的约为 +2kcal/mol。在解折叠变性蛋白质中所有残基都被水包围。这意味着，如果只考虑疏水作用，则在天然球体内部含 Ser 的蛋白质结构比同样在内部含 Leu 的结构更稳定（小于 2kcal/mol）。因而，粗略地说，能够使核心含 Ser 的蛋白质稳定的所有序列，都能使核心含 Leu 的蛋白质稳定化，但是，能够使内部含 Leu 的蛋白质稳定的序列并**不**能使同样位点含 Ser 的蛋白质稳定化。

如果把该蛋白质结构中更稳定的元件（“内部的亮氨酸”）替换为不那么稳定的元

件("内部的丝氨酸"),能使天然蛋白质结构稳定的序列的数目将如何减少?

让我们正式地讨论这个问题,并试着理解,所观察到的各种不同元件与它们的自由能之间的近似指数依赖性来自何处,且式(16.3)中的温度 T_C 的意义是什么。

我事先向大家表示歉意,因为所有的计算我将都只给出最简化的形式,这意味着它们并不是完全精确的。我的目标是让大家认识问题的本质,而不是领入到数学迷宫中,作者之一(A. V. Finkelstein)曾经与 A. M. Gutin 和 A. Ya. Badretdinov 在里面尽情漫步……

■ 为此,假设 $\Delta\varepsilon$ 为在所讨论的元件在天然蛋白质中的自由能(其中包括它与剩余链的相互作用),如亮氨酸或丝氨酸在天然蛋白质球中心的自由能,减去它在变性蛋白质中的自由能。假设 ΔF 为整个剩余链在所讨论的空间结构中的自由能扣除在变性蛋白质中的自由能。则意味着,$\Delta F+\Delta\varepsilon$ 为蛋白质的天然形态与变性形态的自由能之差。

为了让天然蛋白质稳定,$\Delta F+\Delta\varepsilon$ 应该小于 0,即对于稳定蛋白质有:

$$\Delta F < \Delta\varepsilon \tag{16.4}$$

ΔF 和 $\Delta\varepsilon$ 的值取决于肽链的氨基酸序列。让我们来看大量能维持 $\Delta\varepsilon$ 数值不变的序列[即所有序列在肽链的给定位置上的 Leu(或 Ser)不变;而且该位置处于天然蛋白质的核心内,如果变性蛋白质解折叠,对 Leu 有 $\Delta\varepsilon=2\text{kcal/mol}$,对 Ser 大约为 0,同时对所有表面的残基 $\Delta\varepsilon$ 接近为 0。]。而 ΔF 随序列不同而变化。

在这种情况下,$\Delta F<-\Delta\varepsilon$ 的概率 P^* 为

$$P^*(\Delta F < -\Delta\varepsilon) = \int_{-\infty}^{-\Delta\varepsilon} P(\Delta F)\,\mathrm{d}(\Delta F) \tag{16.5}$$

式中,$P(\Delta F)$ 为给定 ΔF 值在随机选择的序列中碰巧出现的概率。

ΔF 值由大量各种相互作用的自由能和固定大量各种氨基酸残基的熵组成。因此加起来的"随机"值有很多,于是,根据数学统计上的"**中心极限定理**",ΔF 等于某个值的概率 P 具有简单的、所谓高斯形式:

$$P(\Delta F) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{(\langle\Delta F\rangle - \Delta F)^2}{2\sigma^2}\right) \tag{16.6}$$

式中,$\langle\Delta F\rangle$ 为 ΔF 的平均值(按所有序列进行平均),σ 为 ΔF 对 $\langle\Delta F\rangle$ 的平均偏差(更准确些:按同样的序列进行平均的均方根偏差)。

提醒一下,"中心极限定理"描述了大量随机数的总和的期望分布,并回答了"这个总和是这个值或那个值的概率如何?"这个问题。我们的"大量随机数的总和"是 ΔF(它包含在随机选取的,即"随机"序列中的许多相互作用)。于是,数学表明,对于大多数序列,这个 ΔF 总和的值应该在 $\langle\Delta F\rangle-\sigma$ 和 $\langle\Delta F\rangle+\sigma$ 之间,而且所有自由能项目的总和等于 ΔF 的概率很陡峭,在 ΔF 偏离 $\langle\Delta F\rangle$ 时指数性地减小(图 16-4)。而由于绝大多数随机序列显然无法使给定的肽链折叠稳定,则在 $\Delta F\approx 0$

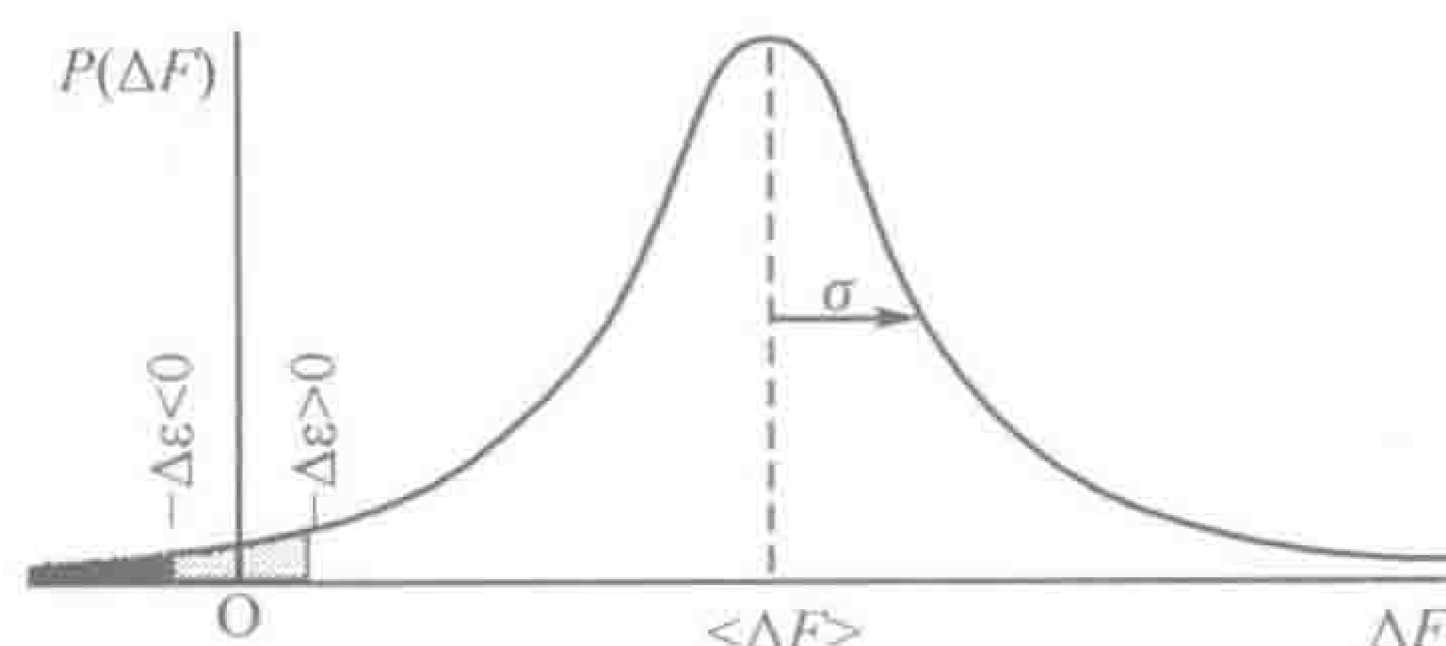


图 16-4　自由能 ΔF 值按各种不同的随机氨基酸序列分布的典型(高斯)形式。ΔF 包含蛋白质链折叠的所有自由能(减去变性蛋白质的自由能)扣除所研究的结构元件的固定自由能 $\Delta\varepsilon$。$\Delta F<-\Delta\varepsilon$ 的值(即满足条件 $\Delta F+\Delta\varepsilon<0$)符合稳定折叠的要求。黑色区域对应于 $\Delta\varepsilon>0$ 时的 $\Delta F<-\Delta\varepsilon$,"灰色+黑色"区域为 $\Delta\varepsilon<0$ 时的 $\Delta F<-\Delta\varepsilon$。第二个区域更大,即如果所研究元件的自由能 $\Delta\varepsilon<0$,则比 $\Delta\varepsilon>0$ 时有更多能使肽链折叠稳定的随机序列

时,P 和 P^* 的值应该在我们感兴趣的"稳定性的边缘"远远小于 1。因此,$\langle\Delta F\rangle$ 值应该不仅为正,而且应该远远大于 σ。

$(\langle\Delta F\rangle-\Delta F)^2$ 的值等于 $\langle\Delta F\rangle^2-2\langle\Delta F\rangle\Delta F+\Delta F^2$,则在 ΔF 为小值(与大的 $\langle\Delta F\rangle$ 相比)时,我们有 $(\langle\Delta F\rangle-\Delta F)^2\approx\langle\Delta F\rangle^2-2\langle\Delta F\rangle\Delta F$,则有

$$P(\Delta F)\approx\left[\frac{1}{\sqrt{2\pi\sigma^2}}\cdot\exp\left(-\frac{\langle\Delta F\rangle^2}{2\sigma^2}\right)\right]\cdot\exp\left(\Delta F\cdot\frac{\langle\Delta F\rangle}{\sigma^2}\right)\tag{16.7}$$

因此[请相信,或者(更难!)请检查],$\Delta F<-\Delta\varepsilon$ 的概率为

$$P^*(\Delta F<-\Delta\varepsilon)=\int_{-\infty}^{-\Delta\varepsilon}P(\Delta F)\,\mathrm{d}(\Delta F)\approx\mathrm{const}\cdot\exp\left(-\frac{\Delta\varepsilon}{\sigma^2/\langle\Delta F\rangle}\right)\tag{16.8}$$

并且常数 $\mathrm{const}=\frac{1}{\sqrt{2\pi\sigma^2}}\cdot\exp\left(-\frac{\langle\Delta F\rangle^2}{2\sigma^2}\right)\cdot\frac{\sigma^2}{\langle\Delta F\rangle}$,是我们不感兴趣的,而我们感兴趣的项 $\exp\left(-\frac{\Delta\varepsilon}{\sigma^2/\langle\Delta F\rangle}\right)$ 表明,蛋白质结构的任意元件的自由能($\Delta\varepsilon$)**指数性地**改变能使包含此元件的肽链折叠稳定的氨基酸随机序列的概率。因此,$\Delta\varepsilon$ 的增加会**指数性地**减少使以能量 $\Delta\varepsilon$ 包含此元件的天然结构稳定的氨基酸序列的数目。

■　补充。上面我们认为,"胜任"蛋白质的是所有那些能够保证 $\Delta F+\Delta\varepsilon<0$ 的序列,即至少对其结构提供最小的稳定性。再过几讲之后我们将会知道,蛋白质结构应该具有一定的"稳定性储备",否则它会在我们的手中熔化,无法进行快速且唯一的自发折叠。因此更正确的是认为,"胜任"蛋白质的序列能保证 $\Delta F+\Delta\varepsilon<-F_{\min}<0$,即 $\Delta F<-\Delta\varepsilon-F_{\min}$。要保证天然蛋白质球的稳定性,$F_{\min}$ 的值只需要几个千卡每摩尔就足够了,即可以认为,$F_{\min}\ll\langle\Delta F\rangle$。因此式(16.8)具有如下形式:

$$P^*(\Delta F < -\Delta\varepsilon - F_{\min}) \approx \text{const} \cdot \exp\left(-\frac{\Delta\varepsilon - F_{\min}}{\sigma^2/\langle \Delta F\rangle}\right) = \text{const}^* \cdot \exp\left(-\frac{\Delta\varepsilon}{\sigma^2/\langle \Delta F\rangle}\right) \tag{16.8a}$$

即(因为在指数项前面的常数是我们不感兴趣的),我们又一次看到了某个元件的能量与该元件在蛋白质结构中被"根据稳定性进行选择"而建立起来的概率之间的关系具有的同样的指数形式。请注意,增强蛋白质的稳定性,即 $F_{\min}$ 值,会指数性地减小对"胜任"它的序列的数目。

■ 所得到的关系式具有指数形式(正因为如此而像玻尔兹曼公式),但是,与玻尔兹曼公式**不同**的是,这里的 $\Delta\varepsilon$ 不是被除以环境温度(更准确些,不是被除以 kT),而是被除以尚不清楚的值 $\sigma^2/\langle \Delta F\rangle$。

这个值有什么意义?首先,我们注意到,$\sigma^2/\langle \Delta F\rangle$ 与蛋白质的大小无关。其实,根据数理统计定律,平均值 $\langle \Delta F\rangle$ 正比于在 ΔF 中相加项的数目(在本例中,大致正比于蛋白质的大小),而对平均值的均方偏差 σ 正比于这些项数的平方根,即 σ^2 也大致上正比于蛋白质的大小。

$\sigma^2/\langle \Delta F\rangle$ 并不随蛋白质的大小而增大,这是非常重要的:这意味着,在式(16.8)中,"缺陷"的能量值 $\Delta\varepsilon$ **不**应该与蛋白质的总能量(正比于蛋白质的大小)进行比较,而是与某种特征能量 $\sigma^2/\langle \Delta F\rangle$ 进行比较,即与某种类似于肽链中每个残基的非共价键相互作用的平均能量(也与蛋白质大小无关!)进行比较。

把这一结果与式(16.8)的指数形式一起考虑,很快就能回答"为什么只有几个 kcal/mol 大小的 $\Delta\varepsilon$ 能对蛋白质结构中的元件的出现频度产生显著影响(如为什么在蛋白质内部的 Leu 比 Ser 要高出一个数量级)"的问题了。发生这种情况是因为当 $\Delta\varepsilon$ 增加 $\sigma^2/\langle \Delta F\rangle$ 时,能使蛋白质天然结构稳定的序列的数目以"e"倍(大约为 3 倍)减小。

因此,$\sigma^2/\langle \Delta F\rangle$ 值是某种类似于肽链增加一个残基(或增加一个自由度)时的能量。或者(提醒一下,kT 是自由度增加时的特征热量),可以假设,$\sigma^2/\langle \Delta F\rangle = kT_C$,其中 T_C 为对于蛋白质链的某种特征温度。是什么样的呢?在"随机"球中只有一个特征温度:该球体最稳定折叠的"冻结"温度。而在蛋白质中有一个特征温度:它熔化(变性)的温度,≈350K(接近于蛋白质的"生活温度",≈350K)。而且可以证明,如果我们只要求蛋白质结构具有最低的稳定性(即如果对其"准随机"一级结构的选择只取决于并不很强的要求),则它大致会在"随机"球凝固的相同温度下熔化(实际上,在略高一点的温度下)。因此可以假设,"构象温度"值 T_C 由蛋白质变性温度所决定。换句话说,可以认为,$kT_C = \sigma^2/\langle \Delta F\rangle$ 等于 0.5~1 kcal/mol。

这不仅可以提议,而且可以证明。"可以证明"这句话意味着相当复杂的理论证明(Shakhnovich-Gutin 理论)我把你们放过去吧。

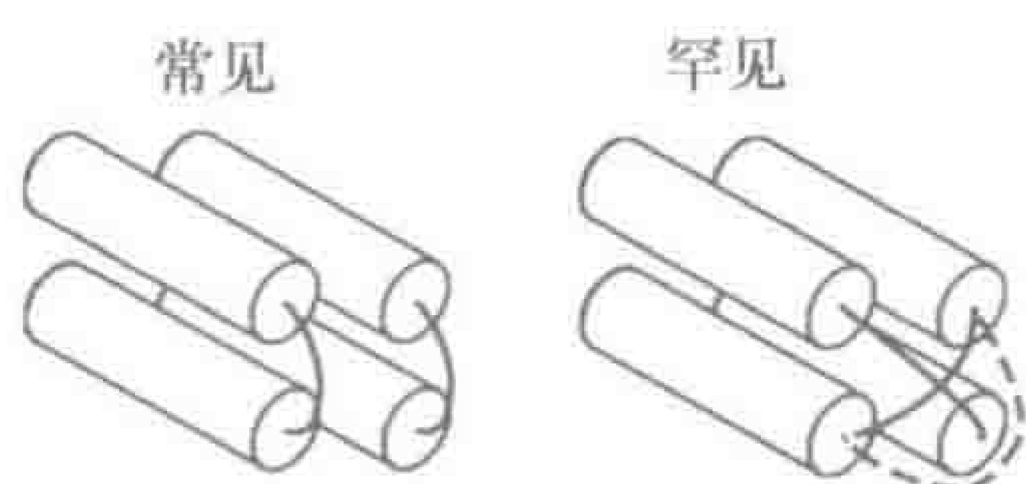


图 16-5　环线的交叠在蛋白质中极少被观察到

■　这样,我们就阐明了,为什么只值几个千卡每摩尔的缺陷(在蛋白质总能量大得多的背景上)可以实际上禁止了很多蛋白质结构模式。因为价值 1kcal/mol 的"缺陷"(任何缺陷)会使对"胜任"蛋白质的序列数目减小到 1/5, 2kcal/mol 的"缺陷"——1/20,如此下去。

因此,尤其是,在蛋白质中极少观察到不规则连接线的交叠(这种缺陷价值 2~3kcal/mol,最大 5kcal/mol)(图 16-5)。

课堂里的声音:不得不打扰一下。您的叙述的整个逻辑的基础在于,无法形成稳定结构的序列被剔除了。完全合乎情理。但那些能产生***超稳定***的结构的序列不会被剔除吗?要知道,所观察到的天然蛋白质的稳定性***总是***并不很高!并且,您也多次提到蛋白质变性温度与该蛋白质在生物体中的生存温度有关。但这样就得出结论,***超稳定***的蛋白质结构[而这意味着(根据您的逻辑!)***超稳定***的结构元件同样如此]应该被剔除。然而您对此完全没有考虑!

讲课人:制造出稳定的蛋白质很困难。对此尤其是在设计人工蛋白方面累积起来的实验可以证明,即能够折叠成某种稳定结构的序列是很难遇到的。而制造"超稳定的"蛋白质就更难了,因为能胜任的氨基酸序列非常罕见!能形成"超稳定"蛋白质(自由能 ΔF 极低)结构的序列数量之少,图 16-4 同样可以证实。因此在保证能形成最低限度可以接受的蛋白质稳定性的肽链中,只有极小部分能够形成"超稳定的"蛋白质……这部分蛋白质(如果选择不是强调"超稳定性",但也不阻止它)***自动地***将会是微不足道的。这对于解释您所注意到的实验事实和关系已经完全足够了……

■　现在,让我们来看第二个问题:为什么熵效应对蛋白质天然结构的稳定性有影响,虽然其中的肽链总是固定的。

这里所有一切也相当简单。

让我们对比肽链折叠的两种模式,如在平行 β 片中的连接线的右旋通路与左旋通路(图 16-6)。

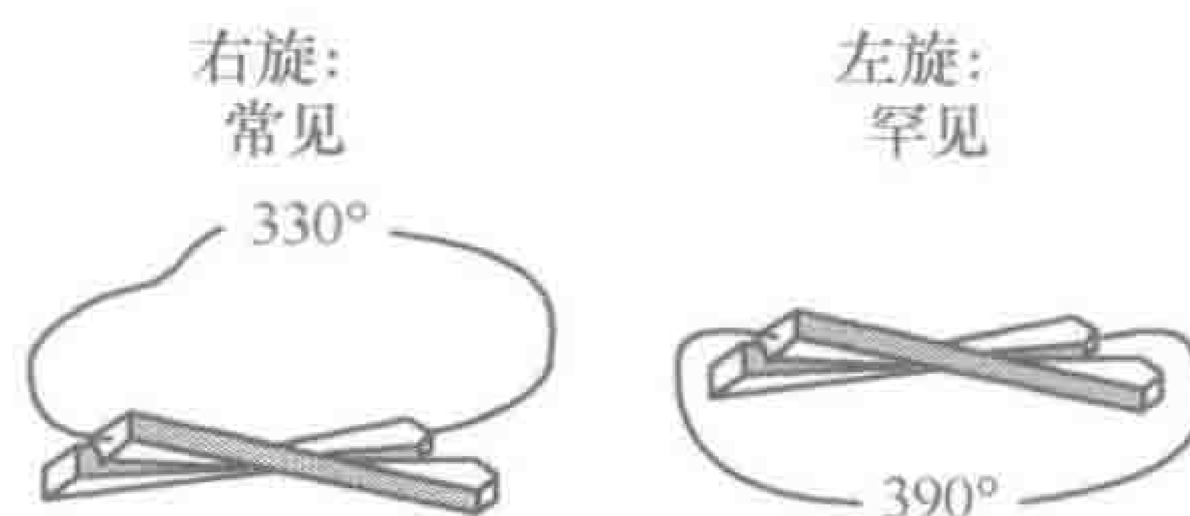


图 16-6　平行 β 片段之间的连接线左旋通路(需要更多的弯曲)在蛋白质中极少观察到,而右旋通路比较常见

由于由 L 氨基酸构成的 β 片倾向于在相邻 β 串之间以大约 −30°角扭转,则***右***旋的连接线需要肽链转动 360°−30°=330°。另一种(***左***旋)连接线转动更多:360°+30° = 390°。高分子聚合物物理学指出,肽链弯曲越强烈,则它的构象越少。这意味着,右旋(弯曲较小)的肽链通路伴随着更多的肽链构象数量,而左旋(弯曲更大,受应力更大),数量不多。换句话说,连接线的"左旋"通路在"左旋"交叉线上会产生"熵性缺陷"。这是可以理解的。然而,这种"缺陷"是如何影响能形成这样或那样稳定连接线通路的序列的数目呢?

随意选择的氨基酸序列可能形成的最稳定的肽链结构，要么是与“右旋”连接线对应的许多种可能构象中的一个，要么是与“左旋”连接线对应的并非很多的构象之一，或者，并不能形成这些构象中的任何最稳定的一种。最后一种，当然是最有可能的，因为在实验中随意熔接的氨基酸共聚物不具有稳定的三维结构。然而，如果只从右旋或左旋这两者中进行挑选，哪种连接线有更大的机会成为最稳定的呢？

每一个独立的含有“右旋”连接线的构象比每一个独立的含有“左旋”连接线的构象既不更好，也不更坏——如果它们具有相同的紧密度、相同的二级结构内容以及其他的话。然而“右旋”的构象更多……因此，我们看到，在我们面前，就像某种以小机会赢取大奖的彩票抽奖（“奖金”是生成稳定的三维结构）：其中“右旋”的连接线有更多的“彩票”（可能的构象），而“左旋”的彩票很少。如果它们最终会中奖的话，哪一个会更容易中奖？当然，是“右旋”的连接线，拥有“彩票”更多的那一种：这两种连接线的胜利概率直接正比于手头的“彩票”（构象）数目……

我们看到，“熵性缺陷”被转化为能量性缺陷：如果在选择时拥有的构象较少，则难以指望有很好的能量。

换句话说，随机氨基酸序列从**大量**“右旋”连接线构象中得到某种稳定构象的机会与从少量的“左旋”连接线构象中得到某种稳定构象的机会之比，就如“右旋”构象数目与“左旋”构象数目之比。或者说：**可能**构象的集合越大（“右旋”连接线的构象更多），则有更多的序列在这个集合中找到自己最稳定的结构。于是每一个这样的序列都从这个集合中挑选出一个最适合于自己的构象。

在球状蛋白中观察到的正是这样：连接线的“右旋”通路很常见，而“左旋”的很罕见。

请回想这些极罕见例外中最显著的一个。那就是左旋 β 棱柱；我们在几讲之前说到过它。它的空间结构确实是独一无二的：在 β 片段之间是连续不断的左旋连接线。而它的一级结构又怎么样呢？看上去，对于球状蛋白来说它同样是独一无二的：它外表看起来完全不是“随机”序列（球状蛋白的典型特征）：它非常有周期性，是由 10 个重复的 18 残基多肽组成的，每个多肽形成左旋超螺旋的一圈（每圈有 3 个 β 串；而在这些串中异常地有很多不扭转 β 片的甘氨酸）。因此：非典型的空间结构是与非典型的一级结构相配合的……

■ 再来看一个与熵性缺陷相关的例子。它已经不涉及肽链的折叠模式，而是涉及球体的分层结构。

α 螺旋穿过中心的蛋白质，和 β 片穿过中心的蛋白质，哪一类应该更经常地被观察到？

由于这两种结构具有相同的与水隔离的表面，它们应该具有大致相同的疏水自由能，**如果**肽链可以插入到那种不容许极性侧链基团处于蛋白质内部的结构中。让我们只来看对随机序列“最困难的”处于蛋白质球最中心的结构片段。在第一种情形下（图 16-7）肽链需要创建内部螺旋；该螺旋的长度取决于球体的直径。为

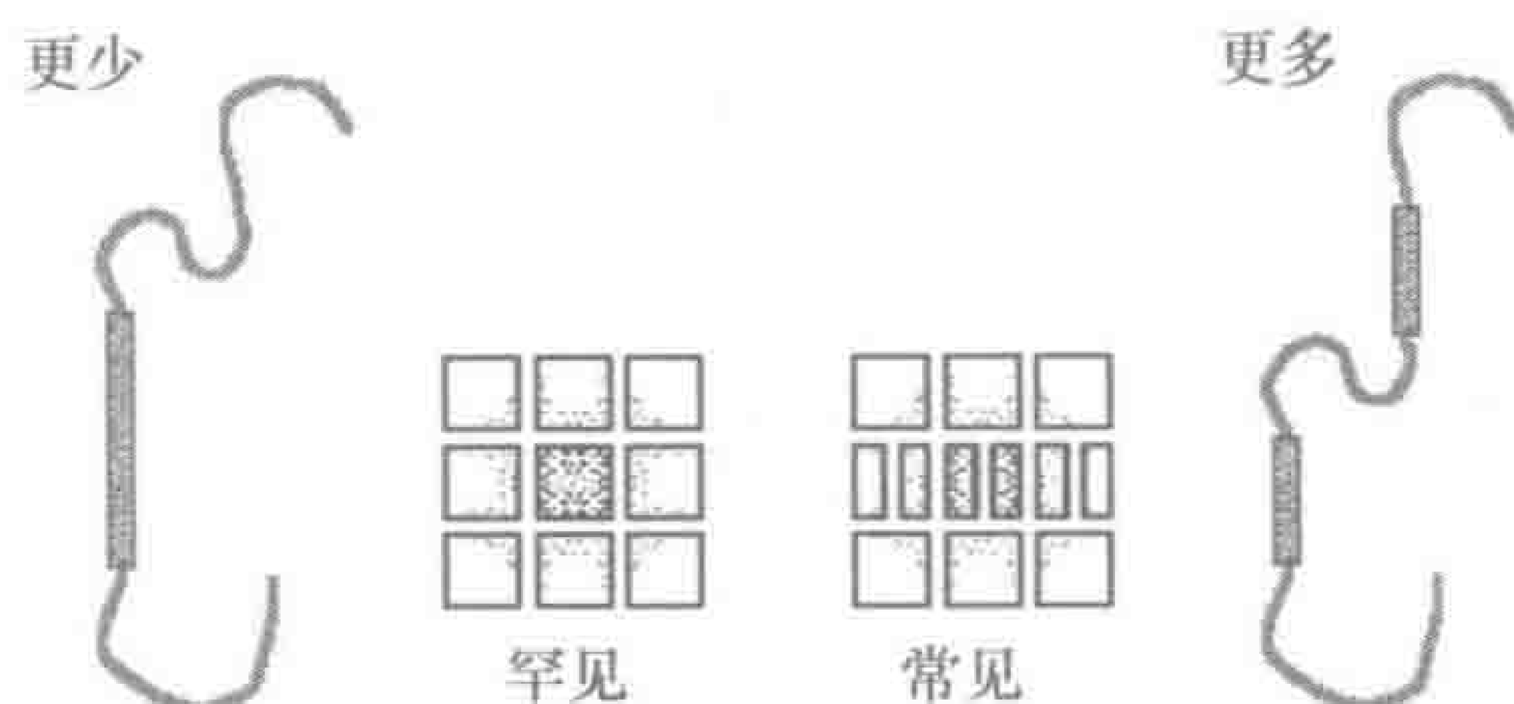


图 16-7 α 螺旋处于中心的多层拼装的可能性,应该比两个 β 段处于中心的多层拼装的可能性要小(事实上观察到的极为罕见)。原因在于,处于球中心的链段应该完全由疏水残基组成,而且这样的结构段的长度被球体的直径所限定,而在相同的长度下,组成 α 螺旋的残基数比组成伸直 β 串的残基数多一倍;因此为了生成内部 α 螺旋(由疏水基团组成的一个更大的模块)需要比生成两个内部 β 段(两个一半大的**以某种方式**排列在肽链中的疏水模块)更"罕见"的序列。在随机序列的**给定位置**存在一个大模块(由 m 个残基)的概率约为 p^m(p 为肽链中疏水残基的分率);而在随机序列的**两个给定位置**存在两个短一半的模块的概率也相同:$(p^{m/2})\times(p^{m/2})=p^m$。然而,一个模块在由 N 个残基构成的肽链中可能有 N 种方式摆放,而两个模块的摆放方式要多很多:约为 $N\times N/2$,即包含两个短模块的肽链比含有一个长模块的肽链多得多

了创建能稳定于球体中心的螺旋,肽链应该含有一个较长的疏水链段。在后一种情形下它应该含有两个短一半(因为肽链在一个残基上行进约 3Å 的 β 串长度,1. 5Å 的 α 螺旋长度)的疏水链段(它们实际上与它们沿链上的距离无关,只要适合于成为两个内部 β 链段就行)。图 16-7 阐明了,为什么创建一个内部 α 螺旋(由疏水基团构成的单个大模块)需要比创建两个内部 β 片段(两个短一半的疏水模块,它们可以以很多种方式摆放在链中)更"罕见"的序列。

因此,可期望的回答(图 16-7)如下,中心穿过 β 片的球体比中心穿过 α 螺旋的球体,能由数量多得多的一级结构所稳定。

而且,确实,在蛋白质内部的 α 螺旋非常罕见(作为这种最罕见的折叠的著名实例是"绿色荧光蛋白",图 16-8),而内部 β 串是很典型的(如在"Rossman 折叠"中)。

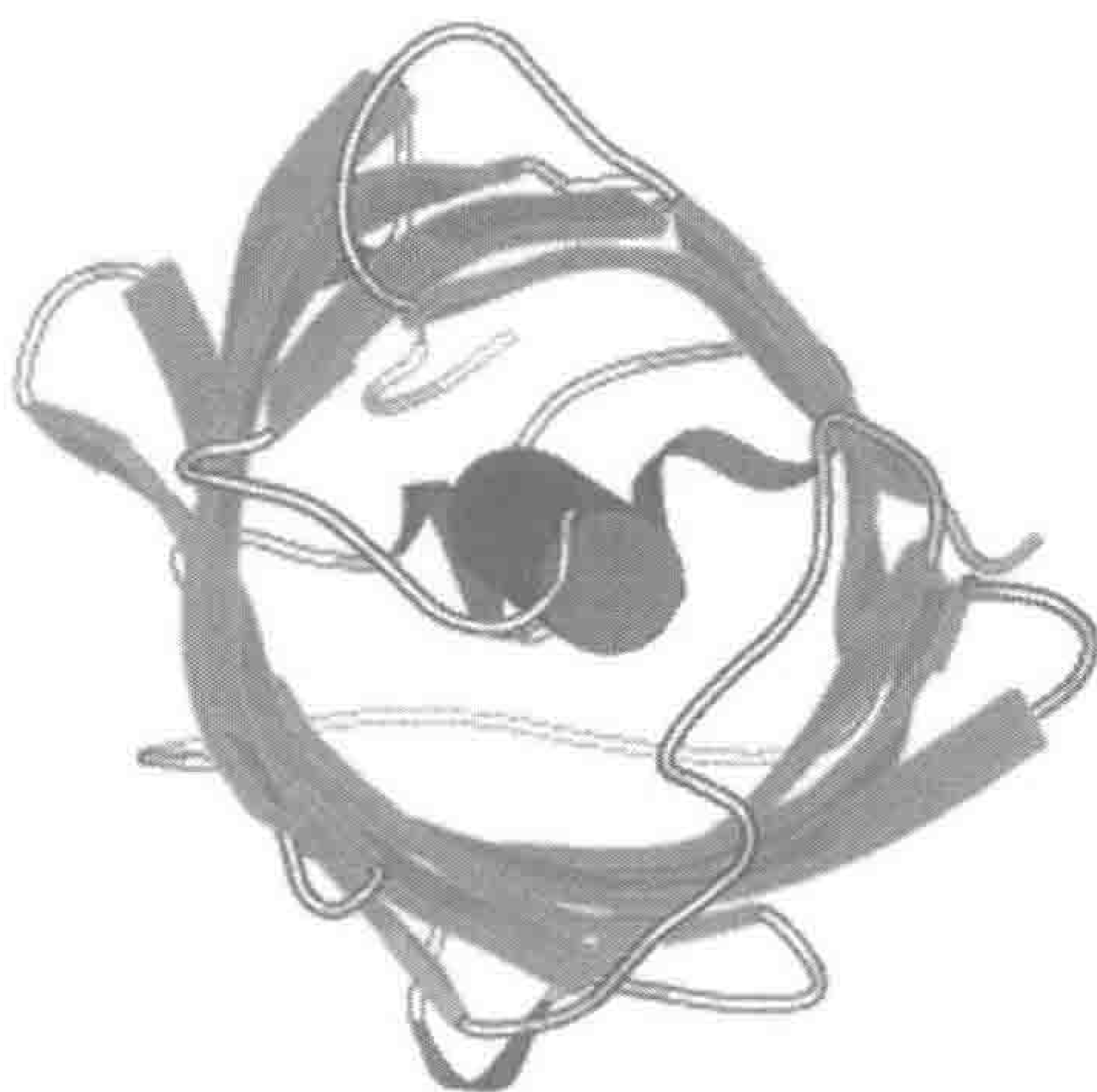


图 16-8 绿色荧光蛋白的异常架构:α 螺旋被 β 片所环绕。中心螺旋实际上不是连续:大约在中间被生色团所折断(该断面未显示于图中)

■ 从同样的观点出发("**给定的结构元件能多么经常地由随机氨基酸序列所稳定?**")可以解释在球状蛋白中观察到的许多其他的规律性。例如,对于含有一定比例的疏水基团与亲水基团的肽链中的域的

平均大小(这个问题曾经被 Bresler 和 Talmud 于 1944 年在苏联的列宁格勒研究过，后来 H. Fischer 在美国研究过)，以及(我们已经讨论过的)α 片段、β 片段和不规则环线的平均长度。

■ 总之，我们的分析表明，某种结构元件的能量越低，它能形成的构象数目越多，则它在稳定的(即所观察到的)**由那些总之能够形成某种稳定球状结构的随机序列构成**的球体中的出现概率，应该越大。

而由于能量和构象数目结合成一个整体构成自由能，所以全然不同的的结构元件在“随机产生的”球体中的观察统计分布应该具有如下形式：

$$出现频度 \sim \exp(-自由能/kT_C) \tag{16.9}$$

式中，T_C 接近于(更精细的分析指出，并不准确地等于，但接近于)由“随机”氨基酸序列形成的球体的凝固温度，而这又接近于蛋白质的变性温度。

而且，实际就是如此，蛋白质的统计分布在总体上具有上述形式：仅为形成稳定球状结构而中选的由随机序列构成的结构的固有统计分布。

■ 再一次提醒一下关系式(16.9)的物理基础。它的结论是，所研究的蛋白质结构元件的自由能**指数性地**改变能够对天然结构中包含该元件的蛋白质给予稳定性的氨基酸序列数目。如果该元件本身是稳定的，则包含它们的蛋白质能够“忍受”很多甚至不利的突变，即有相当多的氨基酸肽链能够使包含该元件的蛋白质稳定。如果该元件是不稳定的，则包含它们的蛋白质需要非常仔细地选择自己的一级结构(其空间结构的稳定性甚至容易被少量的突变所破坏)，而这样“仔细选择的”序列较少。

这里应该强调，所谓“被禁止的”，即在蛋白质中未观察到的(或者极少观察到的)结构在原则上**不是**不可能：它们只是概率很小，即只能很少数量的氨基酸序列产生。

■ 看来，球状蛋白可以相当容易地从氨基酸的随机杂聚物(或者，更准确些：编码随机氨基酸序列的 DNA 片段)中产生。它只需要在少量突变的帮助下使初始的随机多肽的最稳定的空间结构稍加稳定(从而形成它的唯一的被观察到的肽链结构)，并在它的表面“安装”一个活性中心(以便与周围分子进行“生物学上所必需的”相互作用)。另外，也许还需要从蛋白质表面清除那些可能参与“生物学上有害的”缔合作用(如把血红蛋白黏在一起而引起镰刀细胞贫血症)的残基。

课堂里的声音：是否可以这样理解，您提出，具有“新”架构的蛋白质是从随机序列中产生的，而不是从某种“旧”架构通过强烈的蛋白质突变而产生？那些稳定性与许多随机序列相适应的肽链折叠模式是多次从随机序列中产生(在它们的同源家族中有多少个，就有多少次)，而那些只包含一个同源家族的模式只能产生一次？

讲课人:可以足够确信地说的是,不能忽略的一小部分随机序列可以生成蛋白质;而且有一些折叠模式(尤其是那些在蛋白质中常见的那些)与其他"罕见"模式相比,其稳定性与更多的随机序列相适应。至于你说的关于产生来源的问题,那是历史事件,我想现在没办法回答。尤其是,没法说,肽链折叠的"流行"模式的样本是否多次产生。它们有可能是多次从不同的随机序列中产生。也有可能是只是一次产生(并且不是从随机序列中,而是从某种其他蛋白质的小块中产生),然后在进化过程中,具有共同根源的序列在保存空间模式的框架内非常广泛地分化(因为"流行的"模式能够与极不相同的序列兼容),以致这些蛋白质的同源性和遗传关系的所有痕迹都磨灭了,使我们完全无法辨认。我只想强调,"流行的"、与大量序列相兼容的结构模式要比"罕见的"结构模式能对蛋白质的任何起源场景和后续进化适应提供更大的空间。

■ 在这个关节点不能不提到,当代的蛋白质全新(*de novo*)设计工作中广泛使用了氨基酸序列大量随机突变,当然要伴随着对"合适"(如与某些东西相结合的)突变体的选择,也使用了寡肽随机混洗(也加上同样的选择)。选择"合适"随机序列的当代方法(我只提到也许是最强大的方法的名称:噬菌体展示技术。对此请到分子生物学教科书中去阅读!)能够分析大约 10^{10} 或更多的随机序列。在它们中间有时能找到"蛋白质样"产物,这表明,在由几十个氨基酸组成的所有随机多肽中,这些"蛋白质样"的肽链的存在比率大约为 10^{-10}。

我特别要强调,蛋白质产生于随机序列可能只能认为是对于球状蛋白而言,因为正是它们的序列在外表上很像于"随机的"(即最大量的)共聚物(图 16-2)。与此同时,纤维蛋白和膜蛋白的一级结构(以及空间结构)很明显看得出是"**非**随机的"构造(但是顺便说一句,从创建的角度上来说,它们的构造是简单的:纤维蛋白是短模块的重复,膜蛋白是疏水模块和亲水模块的交替)。

■ 上述分析强调了这样一个事实:从物理学观点来看,在进化结果中获得的蛋白质结构看上去非常"合理",就像 DNA 的双螺旋结构和膜双层结构一样。看来,在蛋白质域架构的水平上,进化没有"新发明"任何物理学上不大可能的东西,而是从物理上合理(即稳定的并能快速自组织,对此我们很快会看到)的结构中"进行选择"。这就是蛋白质结构的"物理选择"的意义所在。

蛋白质分子中的协同转变

第 17 讲(两倍)

蛋白质的变性。天然解折叠的蛋白质。协同转变。蛋白质变性的可逆性。球状蛋白的变性是“全或无”型转变。“全或无”型转变的范特霍夫(Van't Hoff)判据。热变性和冷变性。蛋白质分子的相状态图。变性蛋白质看上去怎么样？线团与熔球。熔球的不均匀性。“普通”高分子聚合物球体膨胀时不存在“全或无”型的相变。

前面我们讨论了固定的、“固态”蛋白质结构的稳定性，但是，在特定的外部环境下，蛋白质分子最稳定的结构并不是固态的，而是熔化的、甚至是解折叠的形式。在那些情况下，蛋白质发生“变性”，失去其天然“工作时的”结构。

■ 通常，蛋白质的变性是在体外(*in vitro*)，在特定的异常温度或变性剂[尿素，H^+或 OH^-(即异常 pH)等]的作用下被观察到的。但是，蛋白质的“固态”结构的解体和随后重新自发折叠也可以在活细胞中发生，这在某些过程中，如蛋白质跨膜转运，起着重要的作用。

此外，在生理条件下，也并非所有蛋白质“全靠自己”而具有特定的空间结构。不具有特定空间结构的蛋白质的被称为“天然解折叠的”。有时候是整个蛋白质“天然解折叠”，有时候是蛋白质链的某些部分。某些天然解折叠的蛋白质(或天然解折叠的部分蛋白质链)具有奇怪的、简单的氨基酸组分，而且，可能永远都不具有特定的空间结构。现在仍不清楚，为什么它们对细胞是必需的。有可能是作为某些寡肽的来源。也有可能“不为什么”，就像“自私的 DNA①”一样。然而许多“天然解折叠”的蛋白质一旦与配体或其他蛋白质、DNA 或 RNA

① 基因组研究发现，在多细胞真核生物的基因组中，基因仅是其全部 DNA 序列的一小部分，非基因序列占据绝大部分。非基因序列曾一度被研究者称为“垃圾 DNA”(junk DNA)。这些“垃圾 DNA”中至少有一半是重复序列。过去曾经认为这些重复序列没有功能，只是一些自私的 DNA 序列，热衷于自我数量扩张，而对宿主生物体的繁殖成功没有特定的贡献。但近年来的研究表明，重复序列在生命活动中有着另外的重要作用。

相结合,就会获得自己的唯一空间结构。[有人提出,蛋白质与不稳定结构的结合具有一个优点,即在很大的结合表面上(也意味着,选择性)它并不太牢固。]因此,对蛋白质链的“变性的”(无序的)状态的研究也对某些蛋白质在细胞中的状态有直接的关系。

球状蛋白在体外的变性和复性是深入研究的对象,由于它们与蛋白质的自发折叠问题(即蛋白质链在无穷多的可能结构中如何保持自身的唯一结构)相关,对它们的兴趣仍然持续不断。

在本讲中,我实际上将不涉及蛋白质变性和自发折叠的动力学方面,而专注于这些现象的热力学和结构方面。水溶性球状蛋白是研究得最多的,我也将只讲述它们。

■ 实验给出了什么结果?

业已确认,小蛋白质的变性是协同转变,分子的许多(尽管有时并非全部)特性同时出现急剧的“S 形”变化(图 17-1)。实验曲线的 S 形显示,相应的分子特性从天然蛋白质的特征值变化到变性蛋白质的特征值;而这些 S 形曲线的狭窄性证实了转变的协同性,即它同时包括了许多氨基酸残基。

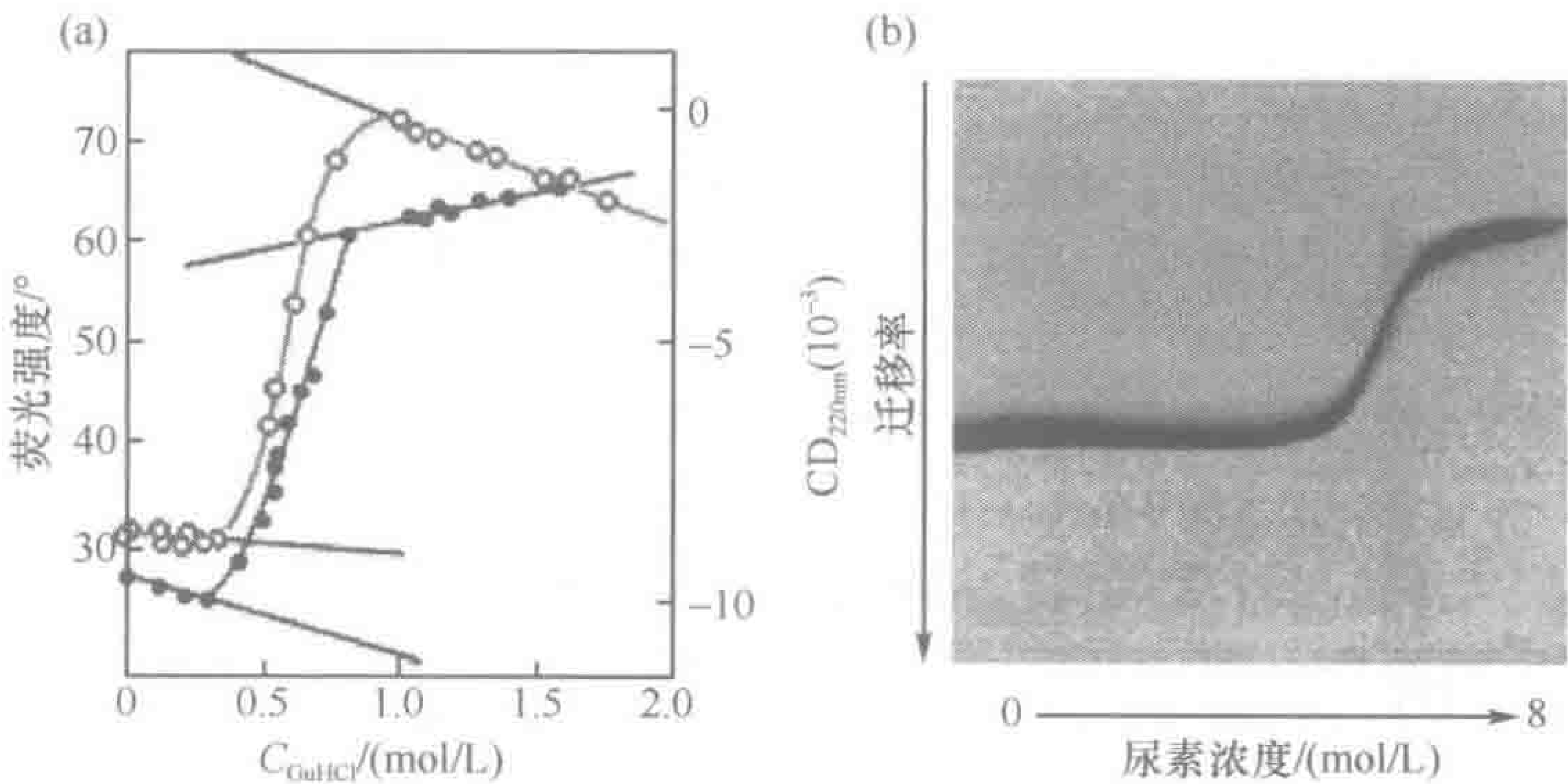


图 17-1 蛋白质的变性伴随着分子的许多特性的急剧的“S 形”变化。在本例中,蛋白质的变性是由变性剂的浓度升高引起的,变性剂浓度以分子在每升溶液中的摩尔数(写为 mol/L 或 M)为单位。(a)磷酸甘油酸激酶在盐酸胍(guanidine hydrochloride)溶液中变性过程中的圆二色性(CD)色谱(波长 220nm)和荧光的同步变化。图片取自 Nojima H. ,et al. 1977. J Mol Biol. 116:429-442,略有修改。(b)细胞色素 c 在不同浓度的尿素中的电泳图:迁移率的滞后是由在变性时分子的膨胀所引起的。图片经许可取自 Creighton T. E. 1979. J Mol Biol. 129:235-264

更进一步,蛋白质变性是以“全或无”型转变发生的(图 17-2)。

“全或无”型转变意味着(你们应该记住),在这种转变中只有初始(天然的)和终末(变性的)状态中才具有可观测的量(图 17-3),而“半变性的”分子实际上并不存在(虽然,当然,它们应该存在,哪怕数量极微,因为一种状态不能借助于“零转运”而转变到另一状态;但“半折叠”的结构只能通过转变内能来反映,对

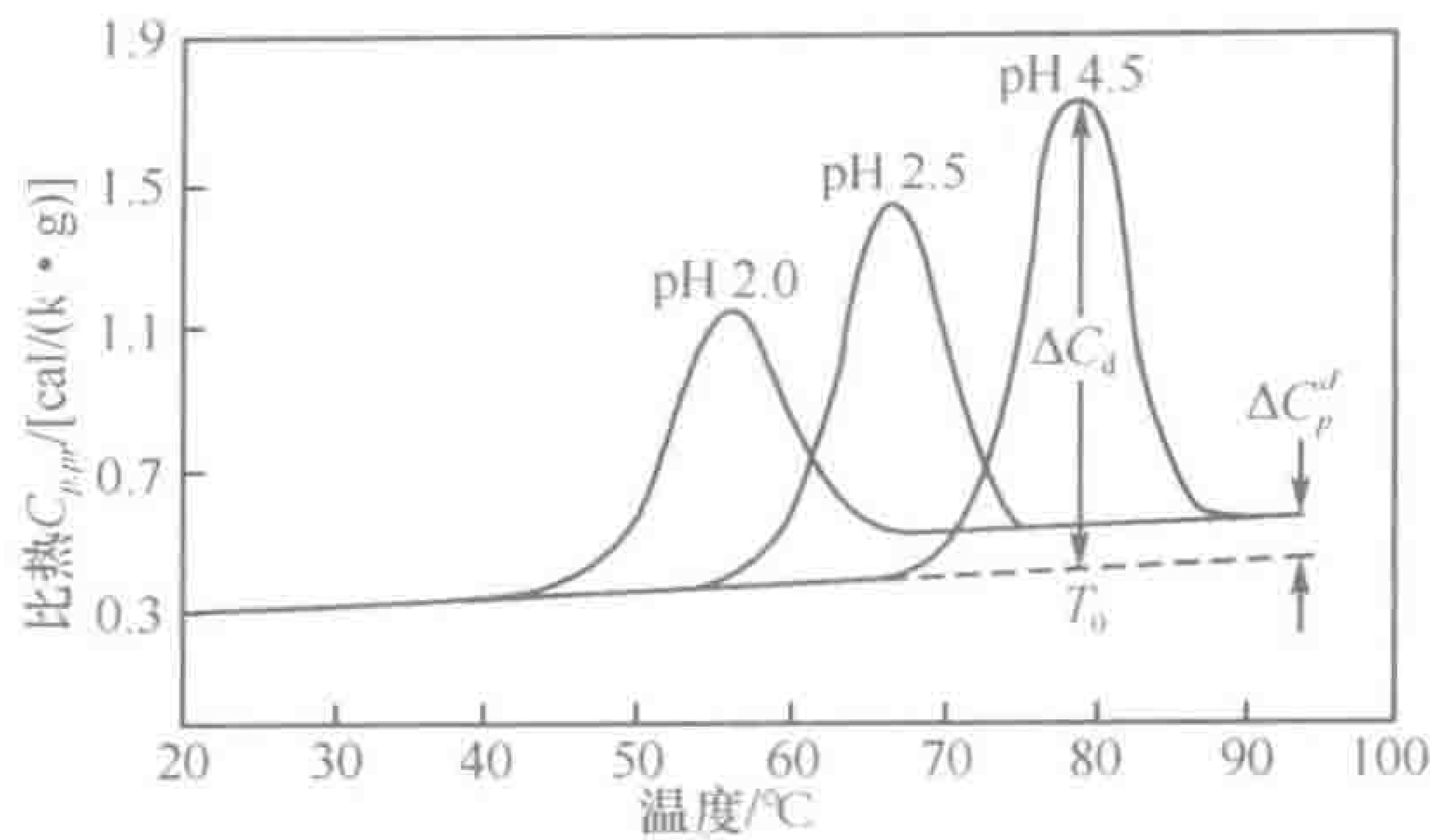


图 17-2 溶菌酶在不同 pH 下的热变性的量热法研究。热容(C_P)的峰值位置确定了转变温度 T_0,峰宽给出了转变温度的宽度 ΔT,峰下的面积确定每克蛋白质熔化时所吸收的热 ΔH。它们($\Delta T, \Delta H, T_0$)满足条件式(17.4)和式(17.5),表明变性是"全或无"型转变发生的。变性蛋白质的热容升高是在蛋白质解折叠(局部地或整体地)后其疏水侧链基团与水的接触表面增大的结果。图片取自 Privalov P. L. ,Khechinashvili N. N. 1974. J Mol Biol. 86:665-684,略有修改

此将在后面讨论)。换句话说,"全或无"转变是宏观系统中的一级相变(如晶体熔化)的微观相似物。然而,与真正的相变不同的是,S 形的"全或无"转变具有非零、有限的宽度,因为参与这种转变的不是宏观系统,而是微观的、极微小的系统。正如实验所显示的,小蛋白质的变性服从"全或无"原则,对此稍后再讨论,而现在最好是明确一下,"全或无"型变性指的是较小的球状蛋白和大蛋白质的单个域,而大球状蛋白的变性,正如大量实验所显示的,是其中各个域的熔化的加和。

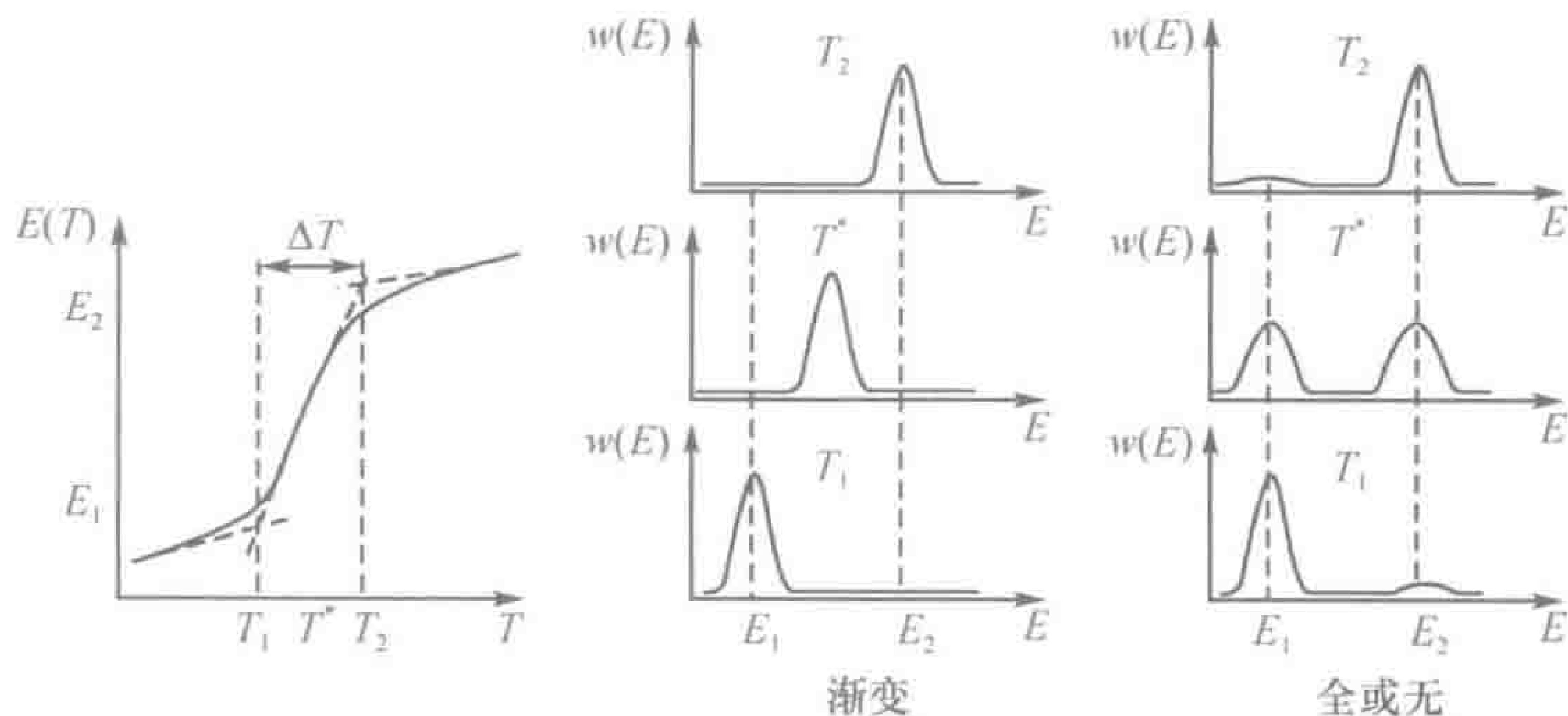


图 17-3 在相同的能量 E(或其他观测参数)与温度的关系曲线形式下,协同("S 形")转变可能是"全或无"型(如蛋白质变性),也可能是渐变型的(如多肽中的螺旋-线团转变)。其区别不在于 $E(T)$ 曲线形状,而在于分子按能量(或按其他观测参数)的分布函数 $w(E)$ 的形状。左图中的虚线说明了转变温度的宽度 ΔT 在图形上是如何测定的

蛋白质变性不仅可以通过加热发生(图 17-2),有时也可以通过冷却而发生(对此将稍后讨论)。此外,蛋白质也可在过高或过低的 pH(引起一种电荷的消失,并从而引起余下的符号相反的另一种电荷之间的排斥)的作用下发生变性。蛋白质也可以随溶液中的变性剂例如尿素(NH_2—CO—NH_2)或盐酸胍($[(NH_2)_3C]^+Cl^-$)的增加而变性(图 17-1),这些物质把水的 O 原

子的氢键拉向自己,破坏水中的氢键平衡,并迫使水分子更积极地扯断在蛋白质中的氢键。蛋白质的变性还可以通过在溶液中添加十二烷基硫酸钠(sodium dodecyl sulfate)分子(它的疏水尾巴能钻入到蛋白质核中而拆毁它)、多种醇和盐等,甚至在特高的压强下发生。

有趣的是,生存在如热-酸-咸泉之类的极端(对我们而言!)条件下的生物体的蛋白质是"亲极端的"(extremophilic):它们只有在那样的极端条件下才能保持为天然形态,而在"生理条件下"(对大部分生物体而言)发生变性。值得指出的是,在亲极端蛋白质的架构中没有任何与"普通的"("嗜常温的")蛋白质明显不同的特性,而对极端条件的适应性是由一级结构的"不明"调整(以增强那些在特定外部条件下能够支持空间结构的稳定性的相互作用)来实现的。不过,讨论这些对我们有点离题太远了……

■ 已变性的蛋白质能不能**复性**并找回自己的天然结构?如果可以,那么它的变性是不是可逆的?

是的(这是从 Anfinsen 在 20 世纪 60 年代的实验之后得知的;顺便说一句,他因此获得了诺贝尔奖[①]),如果蛋白质并不太大而且在体内(*in vivo*)折叠之后**未**经受重大的化学修饰,则被温度、变性剂等造成"轻微"(无肽链断裂)破坏的蛋白质架构在"正常的"环境中能自发地复原。然而,在体外(*in vitro*)有效的变性需要精细地选择实验条件,否则蛋白质可能在溶液中沉淀和/或聚集(对于小蛋白质是分子之间的聚集,而对于大蛋白质,看来还有沿链上距离较远的小块之间的分子内部聚集:已知,随着蛋白质尺寸的增大,复性的能力及其"产出"通常会降低,而实验的复杂性会增大),从而妨碍复性。

蛋白质变性的可逆性极其重要:它说明,形成蛋白质唯一的天然结构所需的全部信息都存在于它的氨基酸序列之中,而蛋白质(严格地说:较小的未被修饰的蛋白质)天然结构本身是平衡的、稳定的;而且这种可逆性使我们可以用热力学来描述和研究这种转变。

■ 蛋白质变性是以"全或无"型转变发生的,这一基本事实是由在俄罗斯科学院蛋白质研究所里工作的 P. L. Privalov 所证明的。他研究了蛋白质的热变性:它通常伴随着很大的热效应,~1kcal/mol 氨基酸残基。

■ 如何证明蛋白质分子的熔化是"全或无"型转变?对此仅给出存在有热容峰(或者,同样的,只给出蛋白质的能量与温度的 S 形关系)是不够的。所有这些只能证明转变的急剧性,说明它同时涉及大量的氨基酸残基,但并没有说明蛋白质是整体地或局部地熔化。要证明蛋白质的熔化是"全或无"型转变,需要比较①从转变宽度中计算得到的"有效转变热"和②该转变的"量热热"(calorimetric heat),即在

---

① Christian B. Anfinsen(1916—1995),美国生物化学家,由于他在核糖核酸酶方面的研究,尤其是氨基酸序列与生物活性构性方面的工作而获得了 1972 年诺贝尔化学奖。1961 年他证明核糖核酸酶在保持其酶活性的同时发生变性后可以重新折叠,从而提出蛋白质采纳其最终构象所需要的全部信息都编码在其一级结构中。

熔化过程中**单个蛋白质分子**吸收的热量的数值。这两个独立数值的重合才能说明，分子是以一个整体熔化的。

这就是关于“全或无”型转变是否存在的范特霍夫(Van't Hoff)判据。由于这一判据的重要性，让我们来仔细地讨论它。

从转变宽度得到的“有效转变热”，即一个独立的“熔化单位”所吸收的热量值。如果等效转变热小于量热热，就说明“熔化单位”小于分子本身，即分子是按部分地熔化的。如果等效转变热大于量热热，则“熔化单位”大于分子，即熔化的不是单个蛋白质分子，而是它们的某种聚集体。

有效转变热与转变宽度是如何联系起来的？我们来分析“熔化单位”，它可能处于两种状态：“固态”，有能量 E 和熵 S；“熔态”，有能量 E' 和熵 S'。为简便起见，假设 E、E'、S 和 S' 与温度 T 无关(对更一般情况的分析，我留给读者……)。

由于状态总共只有两个，则由玻尔兹曼公式容易得出，“熔化单位”处于熔态的概率为

$$P_{\text{molten}} = \frac{\exp\left(-\dfrac{E' - TS'}{kT}\right)}{\exp\left(-\dfrac{E - TS}{kT}\right) + \exp\left(-\dfrac{E' - TS'}{kT}\right)} = \frac{1}{\exp\left(-\dfrac{\Delta E - T\Delta S}{kT}\right) + 1} \tag{17.1}$$

式中，$\Delta E = E' - E$，$\Delta S = S' - S$。“熔化单位”处于固态的概率为(对于“全或无”转变而言)$P_{\text{solid}} = 1 - P_{\text{molten}}$，导数 $\mathrm{d}P_{\text{molten}}/\mathrm{d}T$ 说明 $\mathrm{d}P_{\text{molten}}$随温度变化有多快。简单的计算(通过使用我们已知的公式 $\dfrac{\mathrm{d}(F/T)}{\mathrm{d}T} = \dfrac{\mathrm{d}((E - TS)/T)}{\mathrm{d}T} = -\dfrac{E}{T^2}$)得到

$$\frac{\mathrm{d}P_{\text{molten}}}{\mathrm{d}T} = P_{\text{molten}}(1 - P_{\text{molten}})\frac{\Delta E}{kT^2} \tag{17.2}$$

这里恐怕应该把所有换算更详细地列出来，以便你们不是由于相信而接受，而是通过理解而接受。我们把$(\Delta E - T\Delta S)/kT$ 写为 X。则 $\mathrm{d}X/\mathrm{d}T = -\Delta E/kT^2$，$P_{\text{molten}} = 1/(\mathrm{e}^X + 1)$，$P_{\text{solid}} = 1 - P_{\text{molten}} = \mathrm{e}^X/(\mathrm{e}^X + 1)$，于是有

$$\begin{aligned}\frac{\mathrm{d}P_{\text{molten}}}{\mathrm{d}T} &= \frac{\mathrm{d}\left(\dfrac{1}{\mathrm{e}^X + 1}\right)}{\mathrm{d}T} = -\frac{1}{(\mathrm{e}^X + 1)^2}\cdot\frac{\mathrm{d}\mathrm{e}^X}{\mathrm{d}T} = -\frac{1}{(\mathrm{e}^X + 1)^2}\cdot \mathrm{e}^X \cdot \frac{\mathrm{d}X}{\mathrm{d}T} \\ &= -\frac{1}{(\mathrm{e}^X + 1)^2}\cdot\frac{\mathrm{e}^X}{\mathrm{e}^X + 1}\cdot\frac{\mathrm{d}X}{\mathrm{d}T} = P_{\text{molten}}(1 - P_{\text{molten}})\left(-\frac{\mathrm{d}X}{\mathrm{d}T}\right) \\ &= P_{\text{molten}}(1 - P_{\text{molten}})\frac{\Delta E}{kT^2}\end{aligned}$$

转变中点温度 $T_0 = \Delta E/\Delta S$。在该点 $P_{\text{molten}} = P_{\text{solid}} = 1/2$，而且 $P_{\text{molten}}(1 - P_{\text{molten}})$达到最大值，等于 1/4。如果 $\Delta E/kT \gg 1$，即如果转变发生在狭窄的温度范围内，则

P_{molten}变化最快的点,即导数 dP_{molten}/dT 的最大值也实际上接近于温度 T_0。这里曲线 $P_{molten}(T)$ 的斜率最大且等于

$$\left(\frac{dP_{molten}}{dT}\right)_{T=T_0}=\frac{1}{4}\frac{\Delta E}{kT_0^2} \tag{17.3}$$

在图 17-3 左图中显示了在图形上如何测量转变宽度。把曲线 $P_{molten}(T)$ 的最大斜率切线进行线性外推,直到与对应于天然态($P_{molten}=0$)和变性态($P_{molten}=1$)的基线相交。在转变区间外推的(虚线)$P_{molten}|^{extrap}$值从 0 变化到 1(于是有 $\Delta P_{molten}=1$),而温度变化 ΔT,因此

$$\left(\frac{dP_{molten}}{dT}\right)_{T=T_0}=\frac{\Delta P_{molten}|^{extrap}}{\Delta T}=\frac{1}{\Delta T}$$

即 ΔT 由 P_{molten}值上升最陡的区域所确定,或者,在一般形式下,由能确定分布状态概率的任何实验参数的急剧变化区域的宽度(如蛋白质链的螺旋性急剧变化的区间宽度)所确定。

在最终形式上,"熔化单位"的有效转变热 ΔE 与转变宽度 ΔT 和转变温度 T_0 之间的关系式为 $\frac{1}{\Delta T}=\frac{1}{4}\frac{\Delta E}{kT_0^2}$,或

$$\Delta E=\frac{4kT_0^2}{\Delta T} \tag{17.4}$$

通过用这个方式(**仅由观察到的转变形状**)计算得的"熔化单位"的变性热量 ΔE 与整个蛋白质熔化的量热热(以 $\Delta H/N$ 进行计算,其中 ΔH 为在量热计中所有 N 个蛋白质分子所吸收的热量,而 $N=m/M$,其中 m 为试管中的蛋白质的质量,M 为其分子质量)进行比较。如果

$$\Delta E=\Delta H/N \tag{17.5}$$

则意味着,**整个**蛋白质的熔化是"全或无"型转变。这就是"范特霍夫判据"。

如果 $\Delta E<\Delta H/N$(即如果转变宽度 ΔT 大于这就是范特霍夫判据所需的值),则"熔化单位"小于整个蛋白质,即它是按部分地熔化。如果 $\Delta E>\Delta H/N$,则"熔化单位"大于蛋白质,即作为一个整体熔化的不是单个蛋白质分子,而是某种蛋白质聚集体。

■ 在温度升高时蛋白质结构的破坏:熔化看上去是很自然的;然而还有在温度异常降低情况下的蛋白质"冷"变性(图 17-4)。不过,对于许多蛋白质这种变性没有被观察到,因为在仪器中的水更早就结冰了……

"冷"变性最早是在 Privalov 的实验室被观察到的,该实验室的工作还表明,它(像普通的热变性一样)也是以"全或无"型转变进行的(图 17-5)。

但奇异的是,冷变性发生时就像水的沸腾一样。众所周知,沸腾可以通过两条路径产生:升高温度,或降低内部压强。压缩蛋白质链的疏水压强随温度的降低而急剧减小,于是蛋白质可能"沸腾起来"。

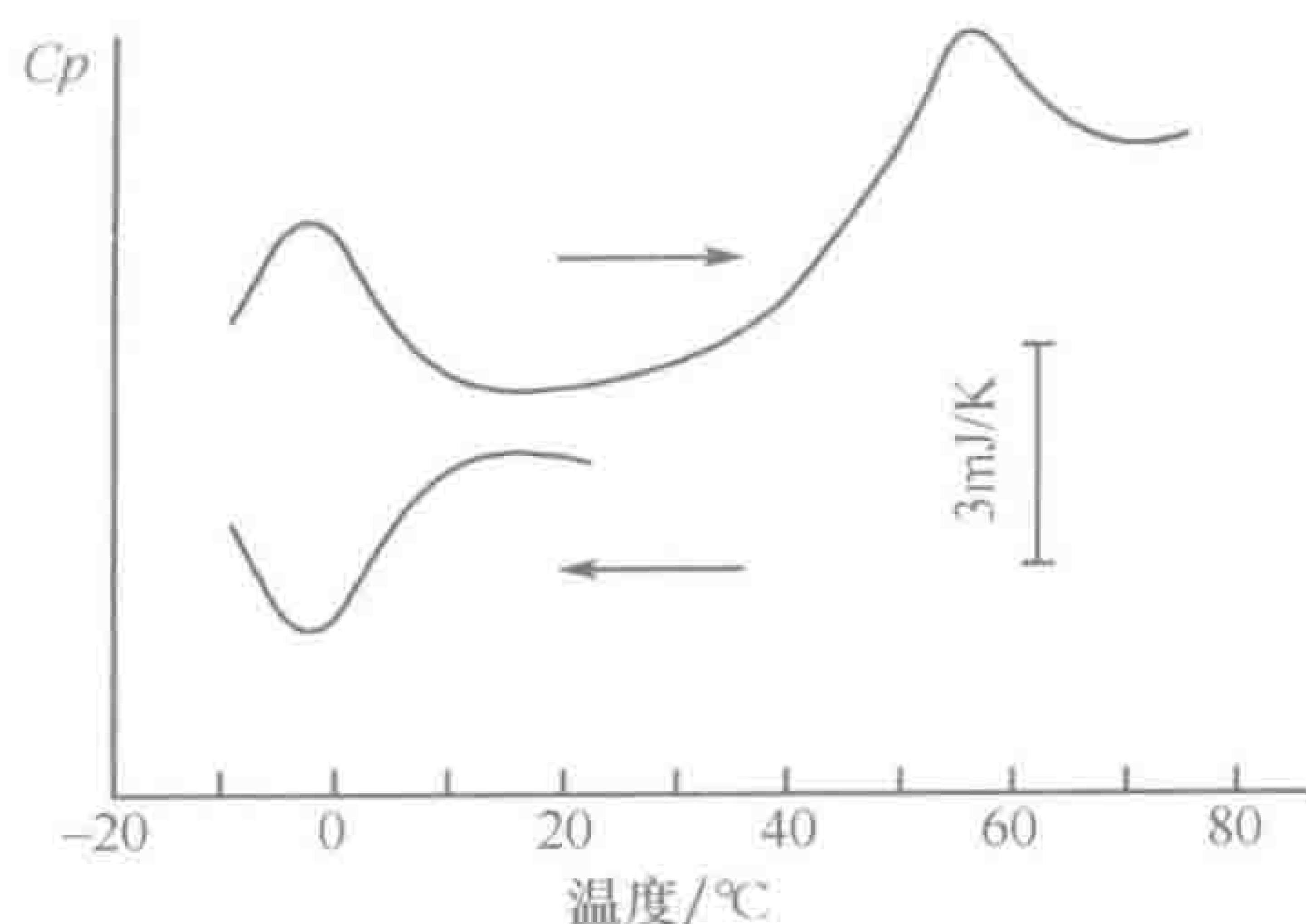


图 17-4 在温度异常降低时蛋白质(脱辅基肌红蛋白,apomyoglobin)的可逆"冷"变性(下方曲线,箭头表示在实验过程中温度降低)。上方曲线表示在温度逆向升高时蛋白质复性(左峰)和它的进一步变性(右峰)。曲线以蛋白质溶液与溶剂的比热容之差表示;未根据溶液中的蛋白质数量进行归一化。图片取自 Griko Yu. V., Privalov P. L., Venyaminov S. Yu., et al. 1988. J Mol Biol. 202:127-138,略有修改

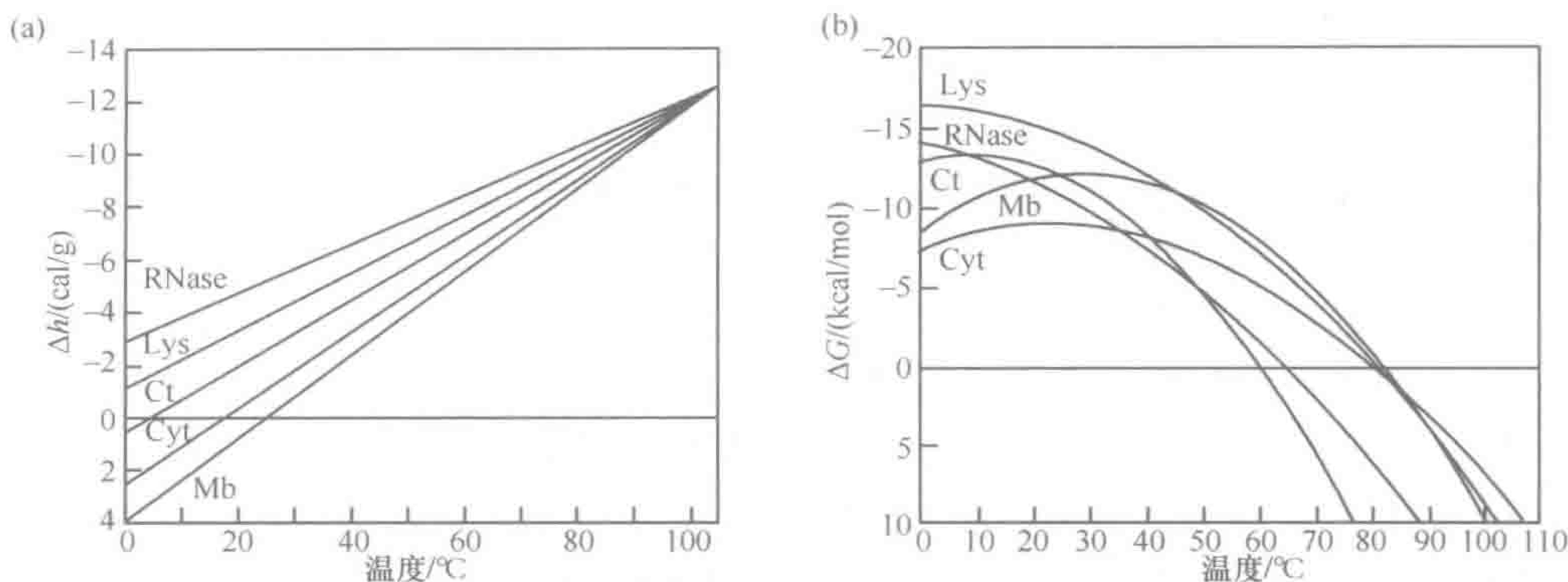


图 17-5 (a)蛋白质天然态和变性态的单位(每克蛋白质)能量差随温度的变化;(b)每摩尔蛋白质分子的自由能变化。图片取自 Privalov P. L., Khechinashvili N. N. 1974. J Mol Biol. 86:665-684,略有修改

顺便说一句,这确实是"沸腾"(意味着自身密度的急剧降低):如果在温度升高时蛋白质能"熔化",则它的体积增大通常不太强烈,而在冷变性时蛋白质链总是完全地解折叠,它所占据的体积会增大很多倍。

■ 这自然给我们带来一个问题:变性蛋白质看上去是什么样子。

对这个问题的激烈讨论还从在 20 世纪 50 年代就开始进行了,此后一直持续了 30 来年,才达成几点一致意见。

问题在于,不同的实验推导出了不同的结论。

确实,大量的**热力学**实验表明,在蛋白质分子的变性状态范围不存在任何协同的结构转变。因此最初认为,变性蛋白质总是非常松散的线团(在如浓尿素溶液之

类的好溶剂中的确是这样）。

然而，对变性蛋白质的非常多的**结构**研究不断地证明，在变性状态的范围内有某些大规模的重组，在分子的完全解折叠的（线团状的）分子形态和天然形态之间存在有某些“中间的”结构状态（图 17-6）。

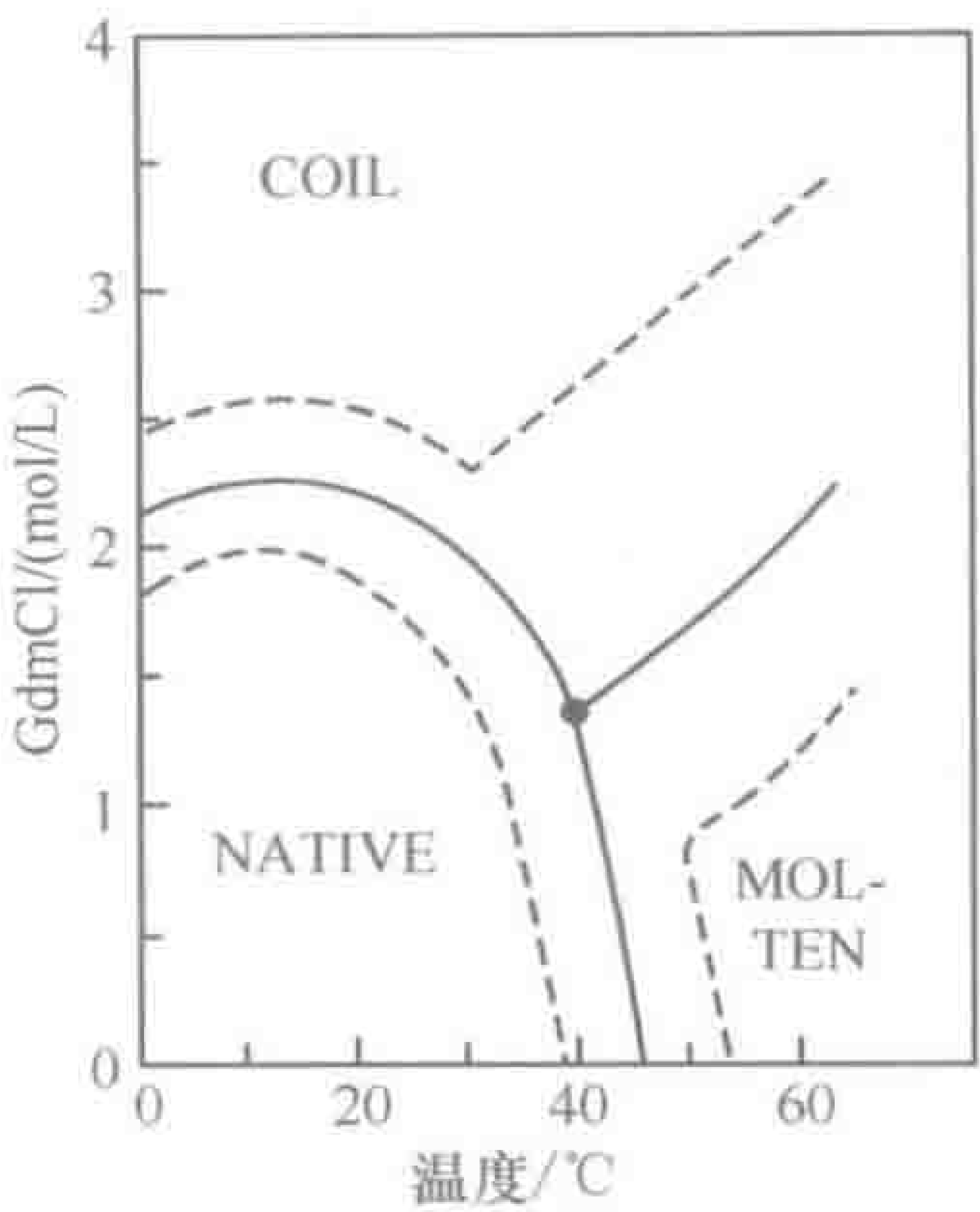


图 17-6　不同温度下溶菌酶（lysozyme）在 pH 为 1.7 的盐酸胍溶液中的构象状态的相图。NATIVE：天然态；COIL：线团；MOLTEN：温度变性态。实线表示转变中间，虚线表示过渡区间（假设由一种状态占优的 9 : 1 关系到另一种状态占优的 1 : 9；可见，与 NATIVE↔COIL 和 NATIVE↔MOLTEN 相比，COIL↔MOLTEN 转变相当宽，“不怎么协同”）。图片取自 Tanford C. 1968. Adv Prot Chem. 23：121-282，略有修改

总之，关于蛋白质变性后的分子状态的实验数据非常矛盾。在某些情况下，蛋白质像是完全“解折叠了”并且没有结构，而另一些相当地有结构和/或紧密的，而需要强烈的溶剂（“变性剂”）才能完全解折叠。

只有通过一系列整套方法来研究变性蛋白质的结构，才能把这个矛盾的图景分析清楚。在这些方法中包括测量黏性（给出蛋白质分子的体积）、远紫外区的圆二色性（CD）色谱（反映蛋白质二级结构）、近紫外区的圆二色性（CD）色谱（反映对称性，即芳香族侧链基团周围的有序性），还有红外、核磁共振（NMR）、蛋白质活性研究以及许多其他方法。在 Oleg B. Ptitsyn 研究组中这一整套武器被应用于蛋白质变性的研究，Alexei V. Finkelstein 作为理论家也参与这个组的工作。

■　实验表明，除了蛋白质活性之外，蛋白质只有两个性质总是在变性时急剧变化。它们是：①蛋白质侧链基团周围的有序性，这被近紫外区圆二色性和核磁共振所观察到［图 17-7(a)］；②球体结构的刚度，这通过借助于核磁共振根据蛋白质极性基团的氢（H）与周围水中的氘（D）的交换速度及蛋白质链水解的加速而被观察到。

蛋白质主链的有序性程度（二级结构），球体大小及其疏水核心的密度实际上既可以保留不变，也可以剧烈变化［图 17-7(b)］。

在这些研究的过程中发现了非常普适的蛋白质解折叠（和折叠）的中间物，后来称为“熔球”（图 17-8，图 17-9）。这种中间体的最重要的性质总结列于表 17-1 中。

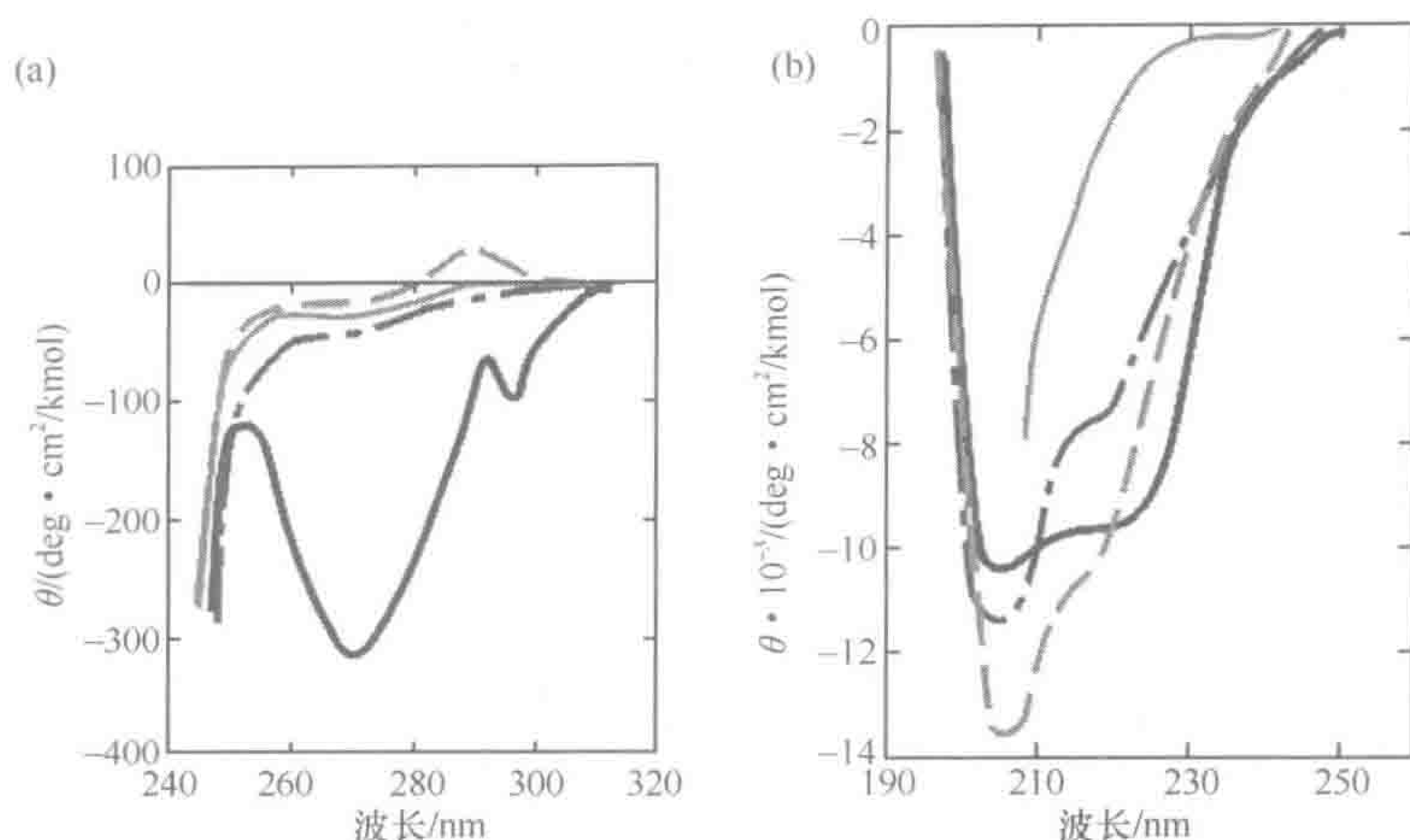


图 17-7 α-乳清蛋白(lactalbumin)在近紫外区(a)和远紫外区(b)的圆二色性(CD)谱:天然态(———)、热变性态(— - —)、酸变性态(— —)和强变性剂变性态(——)。在变性时**近紫外区**圆二色性谱总是强烈地改变,**远紫外区**圆二色性谱实际上只是在由变性剂产生的蛋白质强烈解折叠时才显著改变(见曲线——)。图片经许可取自:Dolgikh D. A., Abaturov L. V., Bolotina I. A., et al. 1985. Compact state of a protein molecule with pronounced small-scale mobility: bovine alpha-lactalbumin. (Figs. 3a, 4). 1985. Eur Biophys J. 13:109-121, Springer-Verlag

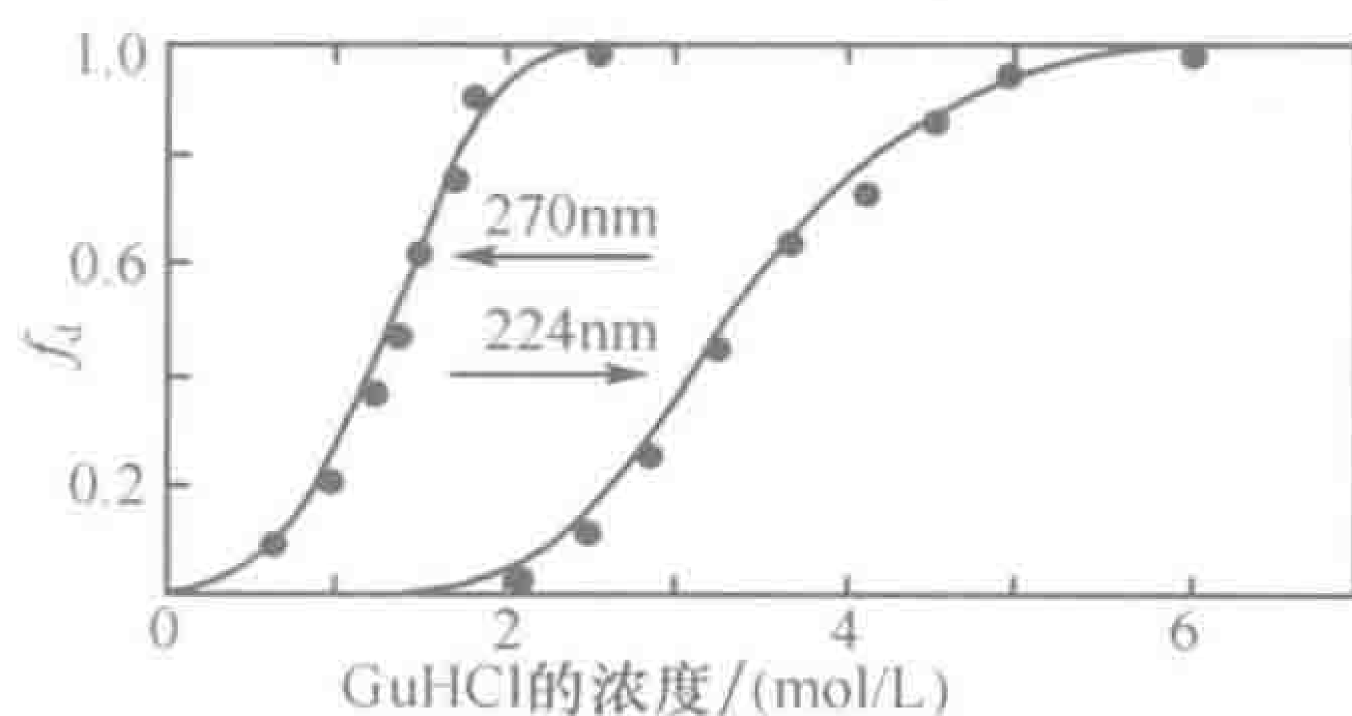


图 17-8 α-乳清蛋白(lactalbumin)在近紫外区(270nm)和远紫外区(224nm)的圆二色性谱随变性剂(盐酸胍)浓度改变的相对变化。$f_d=0$ 对应于天然结构的信号,$f_d=1$ 为完全变性的分子。"**中间态**"存在于第一转变和第二转变之间。在对应于大约 2 mol/L 盐酸胍的区域,所有蛋白质分子都以"**熔球**"的形式存在(后续研究表明如此)。图片取自 Pfeil W., Bychkova V. E., Ptitsyn O. B. 1986. FEBS Letters. 198: 287-291, 略有修改

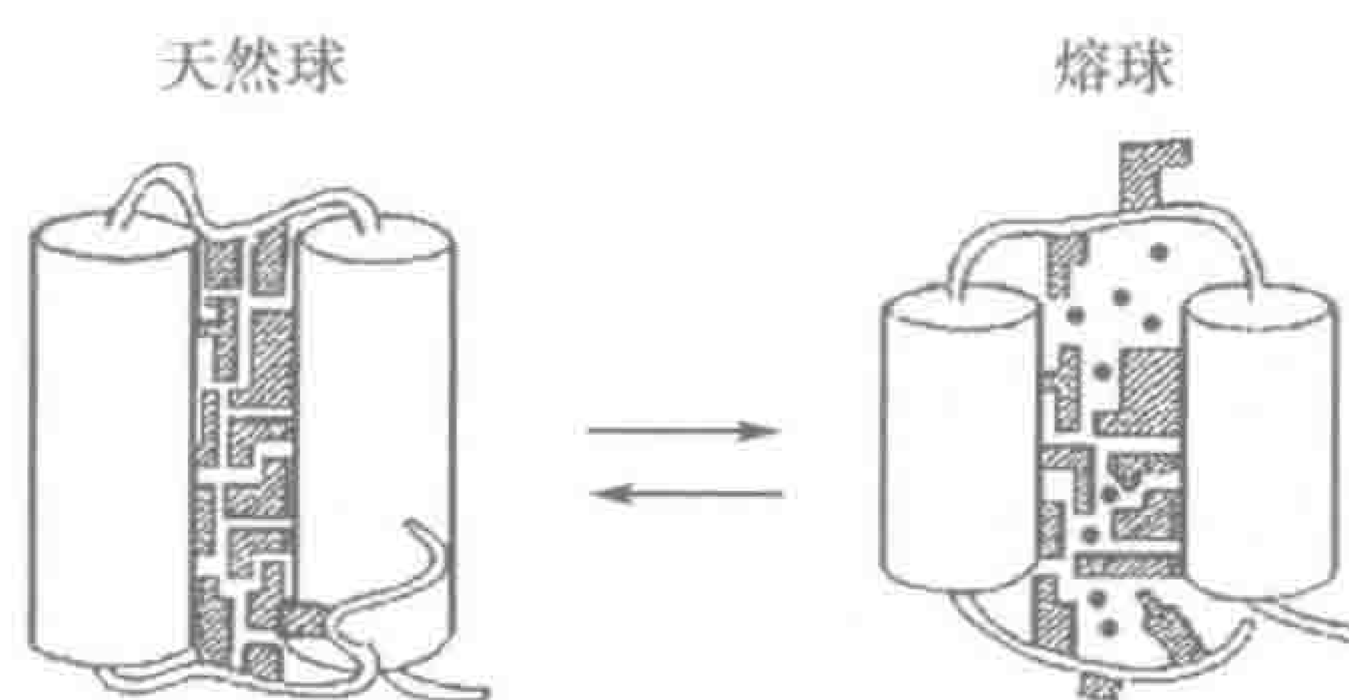


图 17-9 蛋白质分子中的天然球和熔球的示意模型。为简单起见,图中只显示了由环线连接的两个 α 螺旋。每一个二级结构都被大量侧链基团所遮盖;二级结构由氢键所缝合,在插入其中的溶剂"撑开"球体之前是稳定的。如果未添加变性剂,水通常是做不到的。进入蛋白质疏水核心的侧链基团以阴影表示。在熔球中它们获得了运动的自由,得到熵,但是失去密切接触的能量。由于在熔球中出现了孔隙,水(•)渗入其中,但并未拆散熔球

表 17-1　熔球的基本性质及其测量方法

“球体”(与天然蛋白质相似)	“熔球”(与天然蛋白质不相似)
流体力学体积 X 射线中小角度散射 } ⇒紧密性	近紫外圆二色性 ^{1}H 核磁共振谱 } ⇒侧链拼装不唯一
X 射线大角度散射⇒存在有核	H↔D 交换,核磁共振 蛋白水解作用 } ⇒分子涨落
核磁共振(NMR) (自旋回波) ⇒一些芳香环被固定	核磁共振(NMR) (自旋回波) ⇒脂肪族侧链的活动性
远紫外圆二色性 红外光谱 核磁共振+H↔D 交换 } ⇒有二级结构	扫描和等温微量热法⇒无进一步熔化 (通常的规律)
荧光⇒部分 Trp 残基与水隔离	荧光⇒部分 Trp 残基进入到水中
二维核磁共振⇒一些长程接触保持不变	二维核磁共振⇒大部分长程接触解除
色谱法⇒形成“正确的”二硫键的概率	
“熔球”与蛋白质天然态和解折叠态都不同的是:非极性分子的结合增强 (由与蛋白质结合的疏水染料 ANS 的荧光所表明)	

初看上去,熔球的性质是矛盾的。在其中存在有成熟的二级结构,并且实际上通常在数量上与天然蛋白质中相同。可是在其中(与完全解折叠的蛋白质一样)几乎不存在侧链基团的有序性,而那是天然蛋白质的极典型性质。与此同时在熔球中明显保留了侧链基团的天然接触的某些部分,尤其是芳香族侧链基团(但没有脂肪族的),这是核磁共振显示的。它没有天然蛋白质那么紧密(流体动力学体积测量所显示的),而在它的核心却与天然蛋白质几乎同样紧密(X 射线“中角度”散射所显示的),但在熔球中氢交换速度相当高,表明至少单个的溶剂分子能连易地插入到其内部。如果假设熔球具有不均匀性,在其中存在有类似天然态相当紧密的核心和膨胀的环线,在某种程度上可以调和这些矛盾,但是完全天然的核心和完全解折叠的环线的所有假设明显与实验不符,至少与甚至芳香族侧链基团在熔球中也不存在有序环境相矛盾。后者被近紫外圆二色性和核磁共振谱所观察到。

对于极多的(尽管并非全部)的蛋白质,当在天然蛋白质上施加适度的变性影响时,“熔球”出现,只有在添加浓变性剂时才会消失(变成无序的线团)。常常也可以在温度升高时获得类似熔球的状态。而且天然蛋白质转变为熔球总是(当这个问题被弄清之后)“全或无”型转变。“熔球”通常不会在温度继续升高时协同熔化,但它在强变性剂作用下解折叠成线团看上去也是协同的 S 形转变。

然而,有一些天然蛋白质(特别是小蛋白质),直接解折叠成线团而不经过熔球阶段。而另一些蛋白质在某些变性因素(如温度变化或酸)引起的变性中解折叠成熔球,而在另一些变性因素(如尿素)的作用下解折叠成线团。

■　蛋白质变性的结构研究和热力学研究结果之间的矛盾,以及蛋白质的新相态(“熔球”态)的物理本质,需要理论解释。

为了理解熔球态的物理本质,重要的是要考虑到,它是由天然状态通过协同的

变温熔化而得的，而那是一级相变。这意味着，它具有比天然状态大得多的焓和熵，即其分子内部的相互作用被大大削弱了，而蛋白质链的活动性则大急剧增强了。由于蛋白质中大部分自由度与结构的小尺度涨落相关，且首先是与侧链基团的运动相关，正是这些涨落的解放可能使熔球态是热力学有利的。然而小尺度涨落的解放并**不需要**蛋白质链完全解折叠，略有膨胀就够了。不过，在这种情况下，蛋白质中的范德华吸引力急剧减弱：它强烈地依赖于距离，甚至分子尺寸略有增加都足以使它大为减弱（图 17-9）。

总之，所有这一切都与晶体熔化类似：体积的微小增大减弱了其中的部分范德华相互作用，解放了分子的运动。还不清楚的只是，为什么像蛋白质这样如此不均匀的系统，会像由同一种分子（或几种分子）构成的晶体那样，以相变的形式被破坏：要知道，在蛋白质中没有显示出周期性的晶格，而正是晶格的破坏导致晶体熔化。

对“熔球”的研究为理解蛋白质变性的协同性提供了线索。蛋白质变性过程既与高分子聚合物合成过程不同，也与 DNA 和 RNA 不同。

对于后两者，已经知道：它们中没有同样的“全或无”型相变。多肽中的螺旋-线团转变也不是“全或无”型相变。

那么，我们已经讨论过的 β 结构的形成又怎么样呢？因为在那里存在有“全或无”型转变。是的，存在有。然而“非 β 结构”链就是典型的线团，它是完全松散的，其中并没有二级结构。而在蛋白质的变性中经常会形成熔球，其中实际上保留了与天然蛋白质相同数量的二级结构，以及相同的体积。

■ 难道其中不存在二级结构的高分子聚合物[①]的合成不能帮助我们吗？因为在它们中存在有球体-线团转变……也许，这种转变是最类似于蛋白质变性的？有一段时间里人们这样考虑过。

然而，随后弄清楚了，蛋白质变性过程在性质上与“普通”高分子聚合物中的球体-线团转变有本质性的不同。

问题在于，高分子同聚物的实验（极其困难：由于聚集作用的威胁，它们必须在极低的聚合物浓度下进行）、高分子同聚物和统计性高分子杂聚物（不是像蛋白质那样“经挑选的”，而是“普通的”，即“随机的”）的物理理论都与在蛋白质中观察到的并不相符。

它们总是说（图 17-10，图 17-11），高分子聚合物中的紧密球体的破坏应该逐渐开始，与从一相（“球体”）转变到另一相（“线团”）相关联的分子直到最终（或几乎最终）都没有急剧的密度跃变。它们还说，在球体变为线团的时刻之前它已经变得非常松散，而且这两个相的差别主要在于分子密度涨落的尺度（在球体中小，在线团中大），即这种转变完全与一级相变（如熔化或蒸发）**不**相似。而蛋白质变性完全却像固体熔化或升华一样。

---

① 关于高分子聚合物，请参见第 1 讲的脚注。

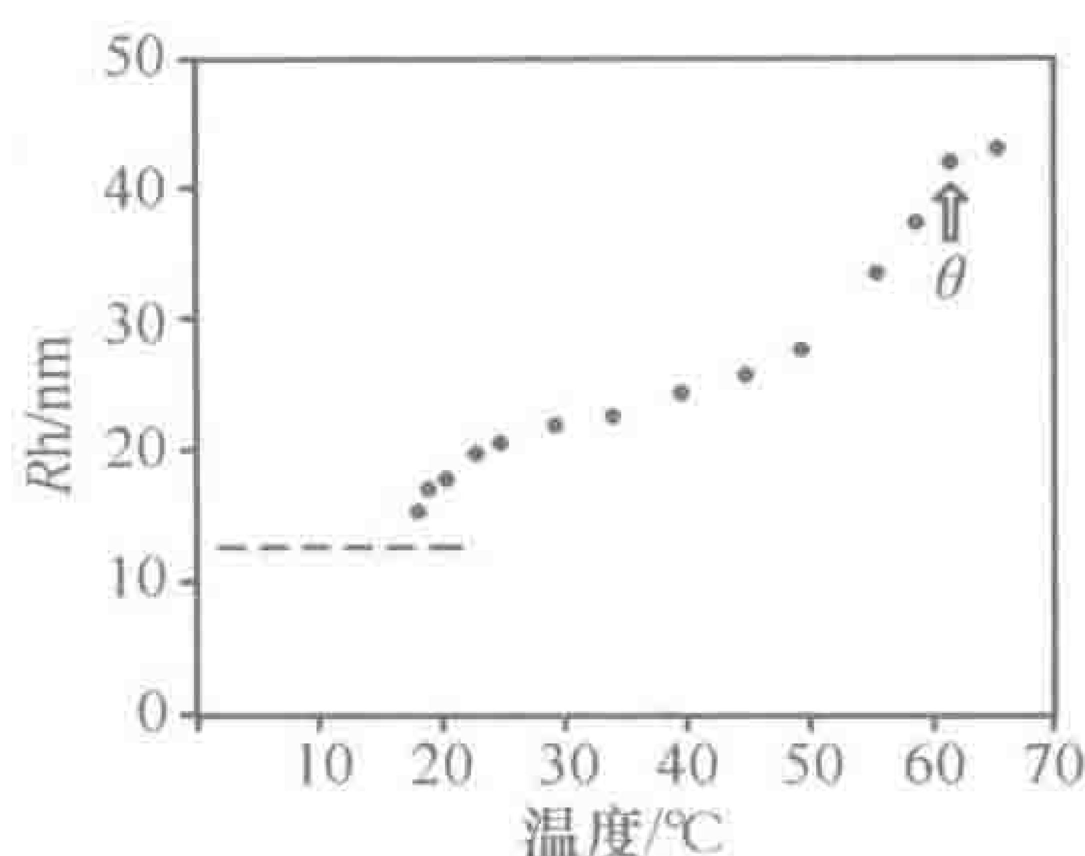


图 17-10　聚甲基丙烯酸甲酯(poly methylmethacrylate)链(相对分子质量 6.5×10^6)在乙酸异戊酯(isoamyl acetate)中的稀溶液的流体力学半径 R_h 与温度的关系(数据来自 Kayaman N.,Guerel E. E.,Baysal B. M.,et al. 1999. Macromolecules. 32,8399)。球体-线团转变点("θ 点")在 61℃(箭头所指)。温度低于 61℃为球体,高于 61℃为线团。计算得的球体半径来自于固态甲基丙烯酸甲酯聚合物(相对分子质量 6.5×10^6),以虚线显示。可见,线团的半径约为固体高分子聚合物构成的球体的 4 倍(即体积增大 2 个数量级,密度变小 2 个数量级),而且在线团→球体转变(温度降低时)之后,链形成了半径为固体聚合物球体 2 倍的膨胀球体(即密度小 1 个数量级)(见 50~30℃区域),在这第一次紧缩之后,还有(在温度进一步降低时)第二次紧缩(见 25~20℃),只有第二次紧缩才生成紧密球体

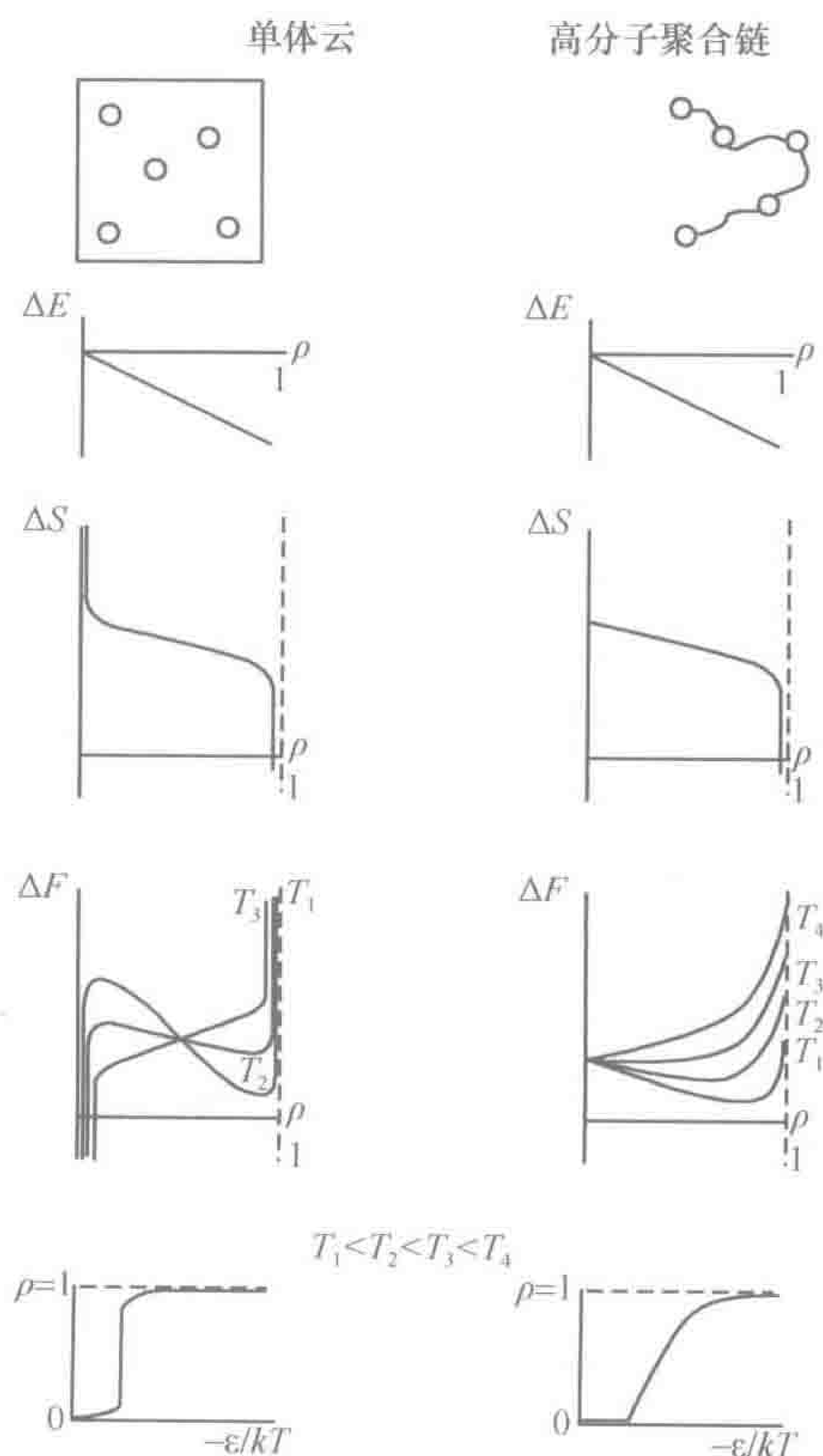


图 17-11　向两个系统("单体云"和"高分子聚合链")增加一个单体的能量 ΔE、熵 ΔS 和自由能 $\Delta F=\Delta E-T\Delta S$ 对密度 ρ 的依赖关系的定性对比。能量图上显示的是相互吸引($\varepsilon<0$)的单体。列出了不同温度 T 下的 ΔF 与 ρ 的关系曲线。在附图(最下方)中显示了两种系统的密度 ρ 与单体相互作用能量(以 kT 为单位表示)之间的定性依赖关系;图中适用于低外部压强条件[在极高的"临界以上的"压强下,松散状态($\rho<0.5$)不再存在]

■ 尽管**高分子同聚物**中的球体—线团转变的经典理论**不**适用于蛋白质变性，我觉得讲述它也并非多余——当然，以最简单的形式。这样的介绍一方面让我稍后能突出那些本质上只能应用于“中选的”蛋白质链、而不能简单地应用于高分子聚合物的东西。另一方面，这一理论完全能应用于已变性的蛋白质分子的行为。最后，我认为了解基本物理模型对通识性教育是必需的。

■ 我建议你们在阅读下面的段落时“用公式检查文字”。当然只读文字更容易，但是它们常常容许不同的解读，因此，如果用公式来检查文字，你们能自己检查自己的理解。

这样，我来讨论最简单的“高分子聚合物”模型：许多(N)个相同的单体(“小珠”)**结合成的链**。并拿它与最简单的、自范德华那个时代(19 世纪！)开始使用的用于描述气体-液体转变的模型“单体云”(即**由 N 个同样的、但未结合成链的单体的整体**)进行比较(图 17-11)。我们可以看到，在高分子聚合物中没有“液态”相、“气态”相分离的类似现象。

事先的注释。下面我将假设，单体互相吸引(否则太无聊了：无法形成稠密相，仅此而已)，而且这种吸引的强度与温度无关，即它只具有纯粹的能量性质。后者经常是不正确的，因为单体在溶剂中是漂动的(回忆一下疏水效应：它随温度而增大！)，但这样更容易找到并分析可能的相变的所有基本事件。再然后可以把任务变复杂(考虑吸引作用的“能量”与温度的依赖关系)并指出什么时候、在什么温度下、这些事件中的哪一些会实际发生。

首先我想给出定性的图片，并解释为什么“单体云”可以经历从稠密相到松散相的跃变，而高分子聚合物不能，即为什么它应该逐渐紧缩。

■ 单体在高分子聚合物中的生活与它在单体云中的生活有什么不同？在聚合物中它不能离开自己在链上的邻居而走得很远；而在单体云中它可以从任何其他单体旁边飞走到想去的地方，只要不撞上容器壁就行。

为什么单体可以形成稠密相：液体(如果单体未以链相连)或球体(如果以链相连)？因为，单体彼此相互吸引，而且，在低温下，当熵的作用降低时(回忆一下：在自由能 $F=E-TS$ 中能量 E 与熵**乘以**温度相比较)，吸引能会打败单体在空间上分散并获得熵的任何企图。总是会打败，除了一种情形之外：当这种分散导致获得实际上无穷大的熵增益。而后者只有当单体云被置于极大的体积中才有可能。则此时这种云的单体不得不在两相之中进行选择：珠滴(能量低，**但同时**熵也小)，或稀薄气体(能量极小，**但因此**熵非常大：能飞到任何地方！)。而中间态能量很低，熵也不高(如当密度为珠滴的 1/2～1/10 时)将具有更高的自由能，因而是不稳定的。因此单体云的稳定状态(在低温下)要么是非常稠密的云相，要么是非常稀薄的云相，但不可能是某种松散的中间相。换句话说，在外部条件改变时(如温度升高)在“单体云”中可以观察到稠密相的急剧汽化。

在“单体云”中的情况就是这样。而在高分子聚合物中呢？**不是**那样，因为在最疏松的球体中单体也不可能远离链上的邻居。即它的熵在最疏松的线团中永远不可能超出某种界限，而且不可能（在低温下）与因球体紧缩而获得的能量相竞争。因此聚合物没有别的选择：在低温下它的单体无论怎么样也不会在整个容器中飞散，它不可能变得那么大，而只能是紧缩成球体，这意味着，与“单体云”不同的是，它不可能经受从稠密相到松散相的跃变。即在外部条件（如温度）改变时只可能出现聚合物球逐渐膨胀（或紧缩），绝不会出现急剧的“汽化”。

现在我们用公式语言把上文重复一遍。

让我们来比较由 N 个单体组成、置于体积 V 中的“单体云”和置于同样体积中、由 N 个单体结合成链的球体。考虑到，“单体云”中和聚合物中的单体都或多或少地按体积分布，我们向每个系统添加另一个单体，并看一看这个单体的自由能（或者，像有些人所说的，它的**化学势**）对系统密度的依赖性。

■　化学势是用于分析相共存的便利工具。要点在于，分子会从化学势较高的相进入到化学势较低的相，这会降低系统的总自由能并使之趋向平衡。而在平衡状态，分子在一相中的化学势等于同类的分子在另一相中的化学势。这意味着，只有当两个系统的不同密度、不同相对应着相同的化学势时，从一相到另一相的跳跃式转变才有可能。而如果系统的化学势随其密度只增加（或只减小），则不可能出现跳跃式相变。

还应该注意到，化学势随密度的增大而**增大**对于相的稳定性是必需的。于是，每一个新粒子进入这个相将越来越困难。这就是稳定的状态：系统越来越（就像弹簧）抵抗施加于其上的作用力。在相反的情形下，系统对其密度的自发涨落不能保持稳定。

总结一下：系统的具体状态取决于具体的外部条件（压强或体积），但系统中存在相变的必不可少的先决条件是化学势的值存在有两个（随密度增大时的）上升图块（分支），由化学势随密度增大而降低的区域所隔开。

■　为了研究化学势的值，让我们首先来看增加到每个系统（高分子聚合物，“单体云”）中的单体的能量，更准确些，是与密度相关（即与以价键结合的单体之间的相互作用无关）的那部分能量。

在两种情形（高分子聚合物和具有相同密度的“单体云”）下，在增加一个单体时，系统的能量都改变**相同**的值。变化量为所增加的单体的相互作用能量值，$\Delta E=\varepsilon\rho$，其中 $\rho=N\omega/V$，为系统密度，即 N 个单体（每个的体积为 ω）在体积 V 中占据的分数，而 ε 为单体在极限压缩系统（在 $\rho=1$ 时）中的所有相互作用。我们将认为 $\varepsilon<0$，即单体是相互吸引的。

于是，能量相同地改变，但是添加到聚合物和单体云中的单体的熵随密度的变化**并不相同**。

在“单体云”中每个单体被赋予体积 V/N。更准确些，由于其余的单体占据体积 $N\omega$，添加到“单体云”中的单体的自由体积为

$$V_1^{\text{cloud}} = \frac{V - N\omega}{N} = \frac{V}{N}(1 - \rho) = \frac{\omega}{\rho}(1 - \rho) \quad (17.6)$$

注意,当密度 ρ 趋向于 0 时,这个体积变得无限大。

在聚合物中不是这样。这里“新”单体只能得到有限体积:

$$V_1^{\text{poly}} = \Omega(1 - \rho) \quad (17.7)$$

而且体积 Ω 受限于单体与链中原有链节的结合,因子$(1-\rho)$说明,如前文所述,其余单体占据整个体积的 ρ 部分。

结果是添加一个单体到“单体云”中的熵变化量为

$$\Delta S^{\text{cloud}} = k\ln[V_1^{\text{cloud}}] = k\ln\left[\frac{\omega}{\rho}(1 - \rho)\right] \quad (17.8)$$

而添加到聚合物链中的熵变化量为

$$\Delta S^{\text{poly}} = k\ln[V_1^{\text{poly}}] = k\ln[\Omega(1 - \rho)] \quad (17.9)$$

这两个函数与密度 ρ 的关系曲线图(图 17-11)在 $\rho\to 1$ 区域相似(在该处 ΔS 中 $\ln(1-\rho)$ 占主导,所以两个 ΔS 在 $\rho\to 1$ 时都减小),但在 $\rho\to 0$ 时**完全不同**。如果 ΔS^{poly} 仍为有限,则 ΔS^{cloud} 不受限制地增大(由于有 ω/ρ 项)。

因此,在低温 T 下,“单体云”的化学势 $\Delta F=\Delta E-T\Delta S$ 值存在有**两个随 ρ 增长的分支**,即有**两个**潜在的稳定状态区域(由于 $-T\Delta S$ 项的作用,在低 ρ 时为气体,由于 ΔE 项的作用,在高 ρ 时为液体)。而在高温下“单体云”的化学势 $\Delta F=\Delta E-T\Delta S$ 值只有一个增长枝,即只有一个稳定状态(气体,由于 $-T\Delta S$ 的作用)。

■ 同时,在高分子聚合物中,$\rho\to 0$ 时熵**不会**变为无穷大,在所有温度下 ΔF 都只存在有一个增长分支。在极低温度下(或者更准确些,在 $-\varepsilon/kT$ 值很大时),这条支路起源于 $\rho\approx 1$,对应于紧密球体状态。然后(随着 T 的升高,更准确些,是随着 $-\varepsilon/kT$ 值的降低)这条分支的起始点逐渐移动到低密度方向(这对应于更松散的球体状态),最后,起始点到达密度 $\rho=0$,并停在那里(图 17-11)。正如你们所见,球体的化学势永远都没有两个增长分支。这意味着,球体膨胀的整个过程永远不会伴有两相的分离。而既然没有两相的分离,就不会有“全或无”型转变,即不会有一级相变。可以证明,在 $\rho\sim 0$(更准确些:在线团密度 $\rho\sim N^{-1/2}$)时会发生二级相变(这个工作是由 I. M. Lifshitz, A. R. Khokhlov 和 A. Yu. Grosberg 以及 De Gennes 所做的),但我们将不去学习这些。

■ 更准确的说明。严格地说,一般性理论认为,比较紧密的球体跳跃式转变为线团只可能在某些很难实现的条件下发生。它仅仅是对那样一些高分子聚合物(其单体成对地相互排斥,但在有大量单体同时相互作用时彼此吸引)是可能的。这种效应已被观察到了,如在非常刚性的链中。然而,与在蛋白质中观察到的不同的是,在这种情形下能量的跳变(以单体计)非常小(如由刚性的细长分子形成液晶时)。能量的大跳变需要非常特殊的单体构造,普通高分子聚合物链的单体不具有

那样的构造。然而,蛋白质链,正如我们将看到的,恰恰具有这样的性质:氨基酸残基只有当以很大的数目参与相互作用时,才开始强烈地相互吸引。为什么是这样,将在下一讲中解释。

■ 因此,普通同质高分子聚合物中的“球体? 线团”转变的传统理论不能解释蛋白质的熔化。(随机异质聚合物中的球体-线团转变理论也不能解释蛋白质熔化,但对它们的讨论超出了本课程的范围。)

传统理论认为,球体逐渐地膨胀,只有在这种膨胀的最后才出现线团,而且完全不是通过**一级**相变。而蛋白质(薛定谔所说的“非周期性晶体”)的变性,是在球体高密度时发生的,经常不能把蛋白质完全转变为线团,而且正如晶体的破坏一样,是一级相变。

■ 对蛋白质变性的这些特性(尤其是它们熔化的相特征)的解释,我们将在下一讲进行。

第 18 讲

为什么球状蛋白的变性是“全或无”型转变？蛋白质核心的紧凑拼装包的解体和侧链基团的解放。溶剂渗入到变性蛋白质中，熔球的破坏，变性蛋白质链随着溶剂力的增强而解折叠。蛋白质链的自然折叠与它的其他球状折叠之间的能隙：蛋白质链与随机高分子同聚物的主要物理差别。“中选的”高分子杂聚物（有能隙）与随机高分子共聚物在熔化时的差别。

我们在上一讲中说明了，经典的“球体-线团”转变理论不能解释蛋白质的熔化。该理论说，球体是逐渐膨胀的，而线团完全不是通过一级相变过程产生的，而蛋白质（薛定谔所说的“非周期性晶体”）的变性发生在球体密度很大的时候，而且恰恰像晶体的破坏一样，是一级相变。

蛋白质的变性也不能用“普通”晶体的熔化来解释：因为在蛋白质中没有周期性晶格，它使晶体区别于液体或玻璃（也是液体，但黏性非常大），而晶格的破裂是晶体熔化的本质。

还令人惊奇的是，蛋白质是异质系统，其中的每一个原子总是处在自己的位置上，而它的变性完全以急剧的一级相变那样进行，类似于在宏观系统中进行的一级相变：而在通常的宏观分子系统中，异质性会使转变显得模糊。

■ 这里有必要注意到一个使“普通”物理学区别于蛋白质物理学以及完全区别于复杂系统物理学的特殊情况。当然，这并不涉及物理定律（很自然，它们在两种情形下都是一样的），而是吸引研究者注意力的东西。“普通”物理学习惯于使用基本上相同的物质：气体、晶体、场等。因此它特别注意在这些“同种物质”中出现的异质性，如准粒子。相反的，蛋白质物理学从一开始就处理具有高度异质性的对象。这里是充满张力的：如果需要的话，可以说蛋白质完全充满了各种各样的涨落，但这也只是心平气和地肯定蛋白质的异质性的一种说法。而围绕这种非同质性系统的整体上的匀质现象，如完全包含整个蛋白质的变性现象，确实是蛋白质物理学特别感兴趣的。

课堂里的声音：蛋白质的熔化真的总是以“全或无”型转变进行吗？即从一种状态到另一种状态之间没有观察到中间体？

讲课人：第一，我现在像平常一样，是讲述单域蛋白质（在大部分多域蛋白质中有多个顺序化的“全或无”型转变，即这些蛋白质可能是“半熔化的”）。第二，在“全或无”型转变中，以及在一级相变中，可以观察到短期存在的中间状态，基本上

是非直接地观察到,根据它们对通过自由能垒(分隔两个非直接观察状态)的转变动力学的影响。对此将在下一讲中介绍。第三,已观察到,在一些小蛋白质中,这种自由能垒是如此之小,以至于其中的转变非常像二级相变,这是 V. Munocha 的结论。而这是在很小的蛋白质中观察到与普遍规律(适用于中等大小的蛋白质和大部分小蛋白质)例外的。

■　要理解蛋白质的变性,就需要解释,为什么蛋白质链存在两种同样稳定的相状态(我们已知,在普通高分子聚合物中这是不可能的),而且为什么它们由自由能垒(我们已知,"全或无"型转变是由它所产生的)所隔开。即需要解释,为什么蛋白质球不像普通高分子聚合物球体那样通过逐渐膨胀而解体。

在解释这些现象时,需要考虑到蛋白质的主要特点(与"简单高分子聚合物"相比):它有具有特殊稳定性的肽链折叠;在蛋白质链中可以活动的侧链基团连接在刚性的主链上;而且天然蛋白质像晶体一样(虽然没有晶体点阵)紧凑地拼装:在蛋白质中,如在分子晶体中一样,原子的范德华体积占据整个空间的 70%~80%,而在液体(熔体)中只有 60%~65%。

因此,蛋白质链的侧链基团能够发生旋转异构化,即从一种允许构象突然跳变到另一种构象。但对这种跳变必须在"跳跃的"侧链基团附近有一定的自由空间。参与构成蛋白质疏水核的侧链基团,连接在刚性的(尤其是由于存在 α 螺旋和 β 片而显出刚性,正如我们看到的,它们对形成核心是必不可少的)主链上(图 18-1)。当溶剂没有进入球体(对此也需要空体积)时,这些结构是稳定的。于是,这些刚性片段是作为一个整体,带着它们所有的侧链基团树枝而运动(至少在球体膨胀之初是如此)。因此,球体的膨胀,即这些片段的裂开(请记住,蛋白质链的天然拼装包具有紧凑拼装成球状、并因此特别稳定的特点),会在每个侧链基团附近形成大致相同的自由空间,而它们要么对于每一个侧链的异构化都不够(当球体膨胀很小时),或者对大量侧链的旋转异构化都足够。因此,只有当微小的膨胀超过某个阀值——"能垒"(图 18-2)时,才能实现侧链基团的解放(以及溶剂的渗入)。这两个事件可以形成蛋白质链的较不紧密、但与其天然态一样稳定的状态,但这只有在超过能垒的密度之后才行。

因此,天然蛋白质球的微小("能垒前")膨胀**总是**不利的,它使其能量升高(因为球体的一些部分已经疏松),但**不**增高其熵,因此**还没有**允许其侧链基团发生旋转异构化。因此球体的自由能在微小膨胀时升高。相反,球体的较大的("能垒后")膨胀允许旋转异构化,并导致(在足够高的温度下)球体自由能的降低。结果是,蛋白质在外部条件改变时的变性不是逐渐进行的,而是跳跃性的(跳过能垒),即按照"全或无"原则进行。

因此,蛋白质能忍受外部条件的变化到某种程度都不发生改变,然后全部马上熔化,像微观固体一样。蛋白质的这种稳定性和硬度进而保证了它在生物体中工作的可靠性。

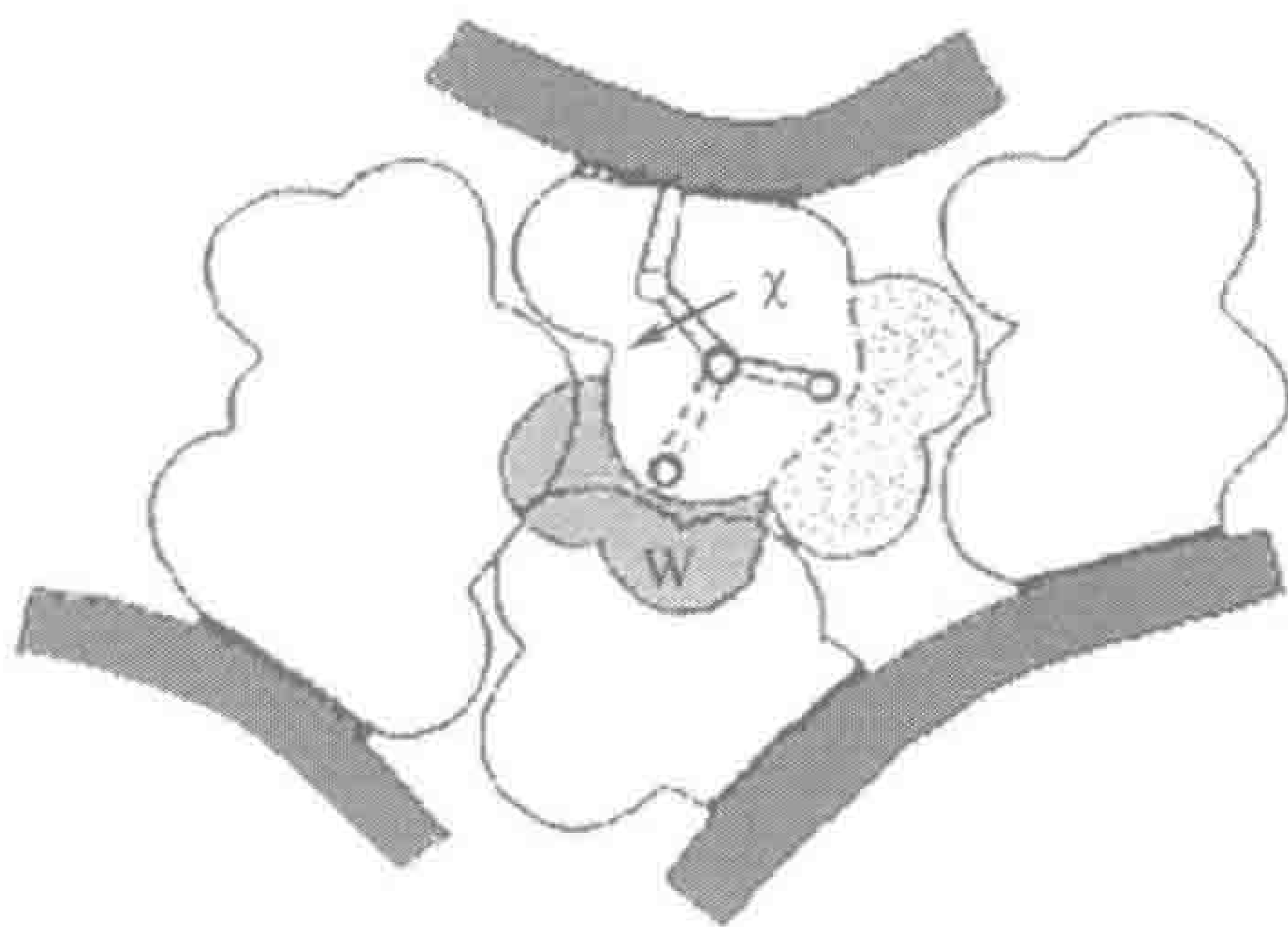


图 18-1　侧链基团拼装的示意图。只显示了核心的一小部分。阴影部分 W 对应于侧链基团(χ 为其旋转角)的非天然旋转异构体；该旋转异构体被紧凑拼装所禁止。它需要额外的空间 $W \approx 30 Å^3$(即≈1/5 的中等氨基酸残基体积)才能出现

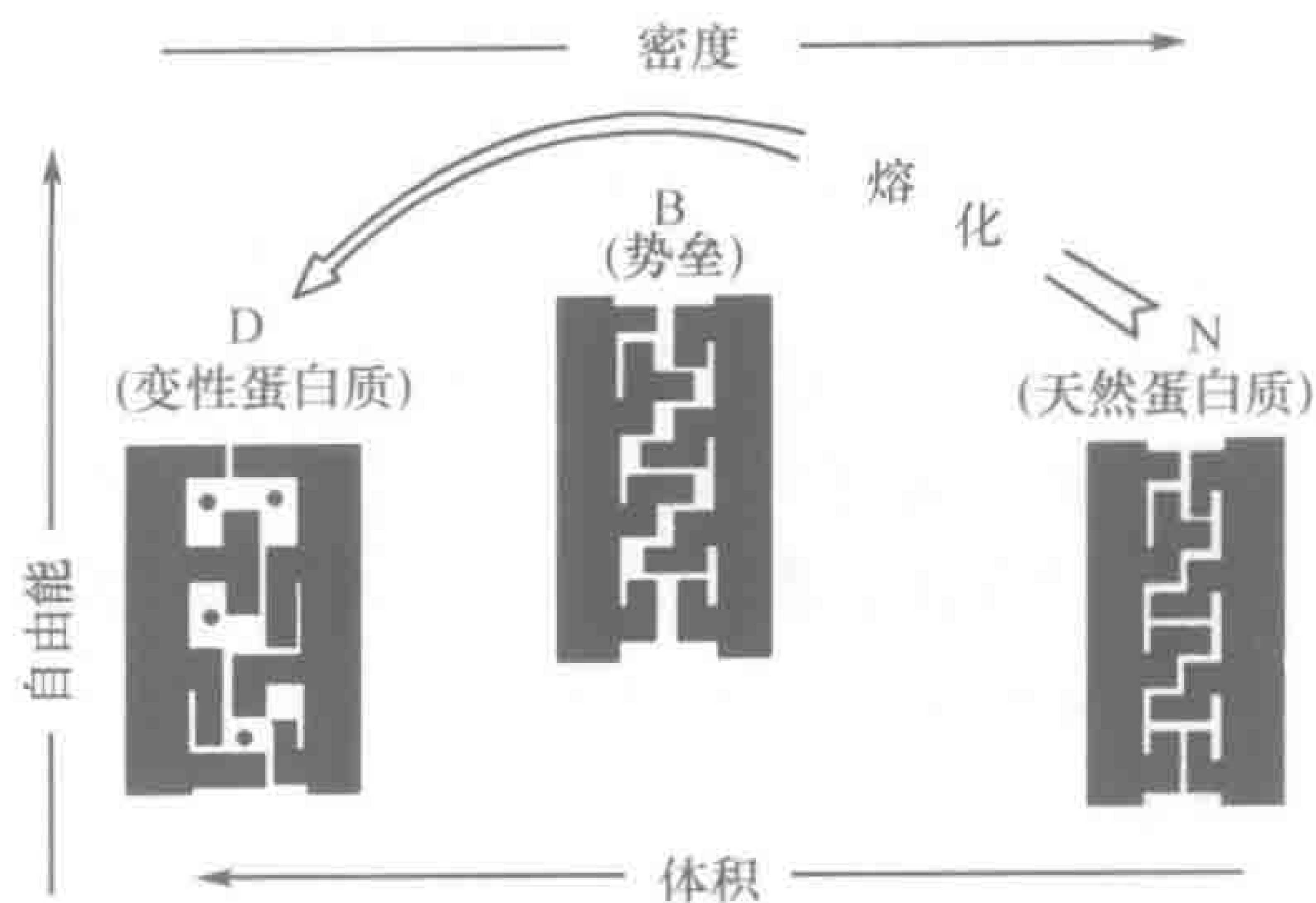


图 18-2　蛋白质的天然态和任意变性态之间存在自由能垒的原因。"能垒"态出现在天然态有微小膨胀之后。在该状态下，蛋白质中的孔隙**已经**导致范德华能量大幅增加，但**还**不允许任何旋转异构体，也不允许水(•)渗入到蛋白质中

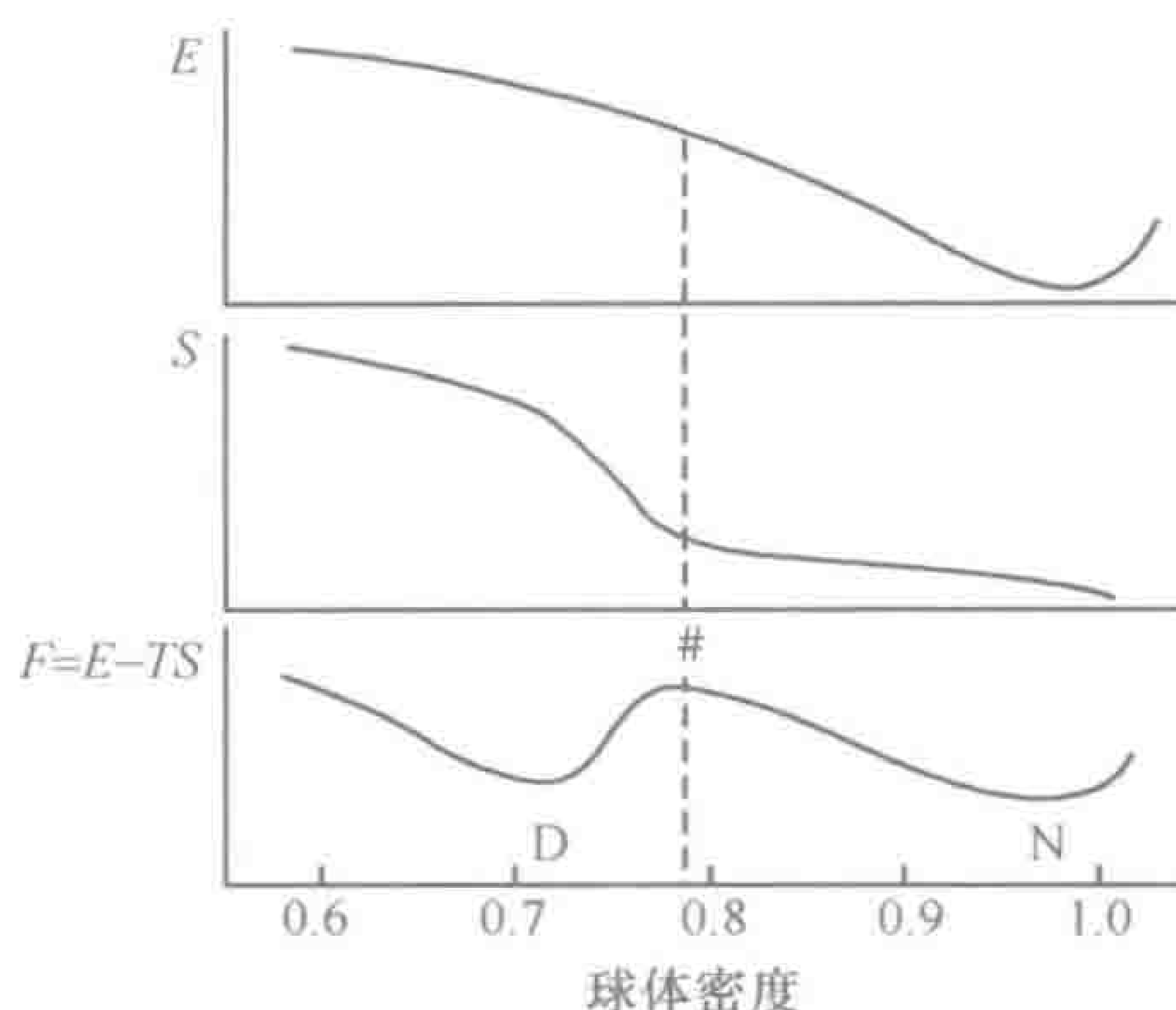


图 18-3　蛋白质变性中"全或无"型转变的原因。能量 E 在球体紧凑拼装(此时密度 $\rho=1$)时最小，并在紧缩或膨胀时单调增大。熵 S 随密度 ρ 的减小而增加，先是缓慢地(在 ρ 为 0.8～1.0 时，这时侧链基团只能振动，还不能旋转异构化)，然后是快速地(这时旋转异构化成为可能)，然后(当旋转异构化已经自由)再一次变慢。熵的不均匀增长与如下效应有关。对侧链基团附近的旋转异构化必须有尺寸至少与 CH_3 基团相当的自由空间。而由于侧链基团与刚性主链相连，这样的空间不可能仅在一个基团上出现，而只能是在许多基团上同时出现。因此就存在球体的特征密度("能垒")，在该处熵开始随球体膨胀而急剧增大。这个密度是相当高的：≈80% 的天然密度，因为 CH_3 基团的体积(特征旋转单位，如图 18-1 所示)约为≈20% 的氨基酸残基体积。结果是，总自由能 $F=E-TS$ 与球体密度的关系曲线要通过一个最大值("能垒"#)，它把天然(紧凑拼装的)球体(N)与更松散的任意变性状态(D)分隔开。由于存在这个能垒，蛋白质的变性是以"全或无"型发生，与变性分子的后续状态无关

换句话说，天然态和变性态之间的相变可以用球体膨胀过程中熵(主要是侧链基团的熵)的跳跃性增长来解释(图 18-3)，而其协同性的相特征是因为侧链基团与主链相连而且不能一个一个地解放。

■　在发现熔球之前,蛋白质变性通常被解释为结构的完全破坏和球体转变为线团。在发现熔球之后才弄清楚,变性的蛋白质可以相当紧密,也可以相当松散,这取决溶剂力和链的疏水性。计算显示,熔球的孔隙(侧链基团自由运动所必不可少的孔隙)被溶剂所充满,因为在热力学上让水分子进入到蛋白质内部孔隙中比那里存在真空要更有利。关于这一点在实验上已证实的是,在蛋白质变性过程中局部体积没有什么改变。但渗入的溶剂如果不被孔隙所强烈吸引(因为它们主要是被微弱地吸引水的非极性基团所包围),则只能填满已经存在的孔隙(侧链基团的自由运动所必需的),而不能产生新的孔隙(对这些运动已经不是必需的了,只是对溶剂是“需要的”),即不能扯破仍然存在于蛋白质中的相互作用力。因此它不能使球体胀大,正如水虽然已经填满了海绵的孔隙,但不能胀大海绵,而变性的蛋白质因此维持为紧密的熔球(而且是“湿的”,饱和地充满溶剂的熔球)。

课堂里的声音:既然您说到“湿的”,那么应该还存在有“干的”熔球?

讲课人:是的。它就是孔隙中未充满水的球体。这样的球体(像“湿的”一样)在理论设想上是由 Alexie V. Finkelstein 和 E. I. Shakhnovich 于 20 世纪 80 年代提出来的,但理论分析和实验数据表明,干熔球的稳定性不如湿熔球,因此较少成为蛋白质熔化的中间体。然而不久前干熔球在蛋白质熔化之前的涨落中被观察到了。(Jha, Udgaonkar. 2009. Proc Natl Acad USA. 106:12289-12294)

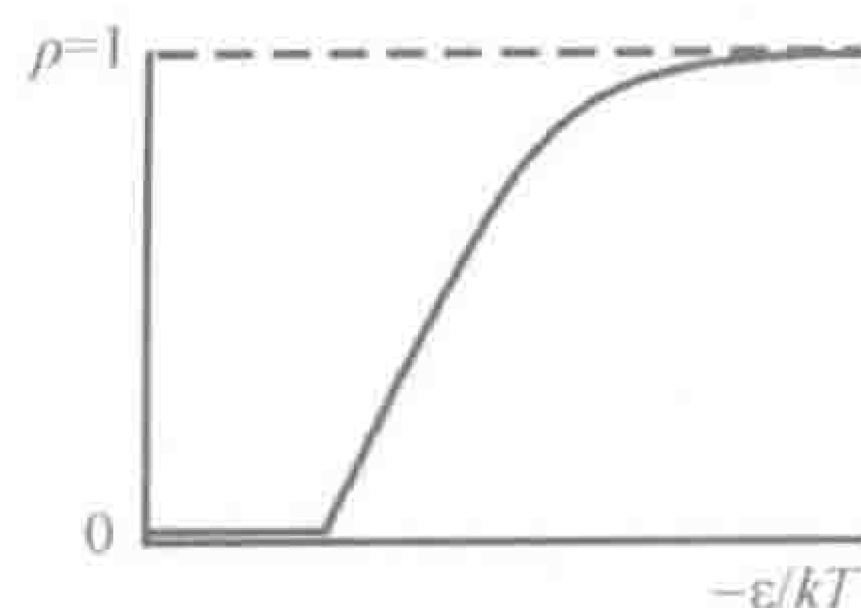
熔球的紧密性是由剩余的疏水相互作用维持的。它们并不很强烈。甚至在大部分链仍具有成熟二级结构和几乎天然折叠拓扑的脱辅基肌红蛋白熔球中(Jennings, Wright. 1993. Science. 262:892-896),剩余的疏水性侧链基因的相互作用至少减弱至天然蛋白质的三分之一(Samatova, et al. 2009. Protein Sci. 18:2152-2159),而在某些这类基团中,这些剩余的相互作用完全不存在了(从而凸显了熔球的不均匀性)。如果剩余的疏水相互作用比较微弱,即如果溶剂被蛋白质中的孔隙强烈地吸引,则它开始扩大孔隙,球体开始膨胀,而且膨胀越强烈,则溶剂与蛋白质链的吸引越强烈,即链节的相互接触就越不利。随着溶剂与蛋白质链的吸引力继续增大,球体膨胀变成无规线团。

在相对较低的温度下,侧链基团解放获得的熵的作用不那么重要,则主要工作落在溶剂上,天然结构直接转变成线团(请回忆冷变性)。在高温时线团可以更快在熔球膨胀之后出现(请回忆 Tanford 的实验相变图;我稍后会再次展示它)。

熔球的膨胀,似乎可以用传统的球体-线团转变理论来描述,如图 18-4 所示:如果蛋白质链的疏水性足够大,热变性蛋白质维持在熔球状态;如果不够大,它会膨胀或(如果溶剂“更好”,或肽链不大疏水)完全解折叠。

这就是最简单的熔球理论所得出的结果。实际上,这一理论认为,熔球并不是准确地从线团方向过来的顺序状态,熔球并没有被“全或无”转变与线团隔开。但这是并没有考虑蛋白质的异质性的简单理论的结果。也许,需要更高级的理论?

图 18-4　普通高分子聚合物的密度 ρ 与单体相互作用能量 ε(以 kT 为单位)的定性关系曲线。这里没有观察到密度的跳变,即这种变化不是“全或无”型转变。然而,从某个 $-\varepsilon/kT$ 值开始,ρ 随 $-\varepsilon/kT$ 的增大速度急剧改变(见曲线拐点),这对二级相变是很典型的。随后这种增长逐渐平息,已经没有任何相变了



■　让我们来看实验得出了什么样的结果。首先,实验显示熔球的解折叠像是协同(“S 形”)转变(图 18-5)。不过,这个“S 形”既与图 18-4(解折叠时密度迅速变化,但没有跳变)相容,也与“全或无”型转变相容。

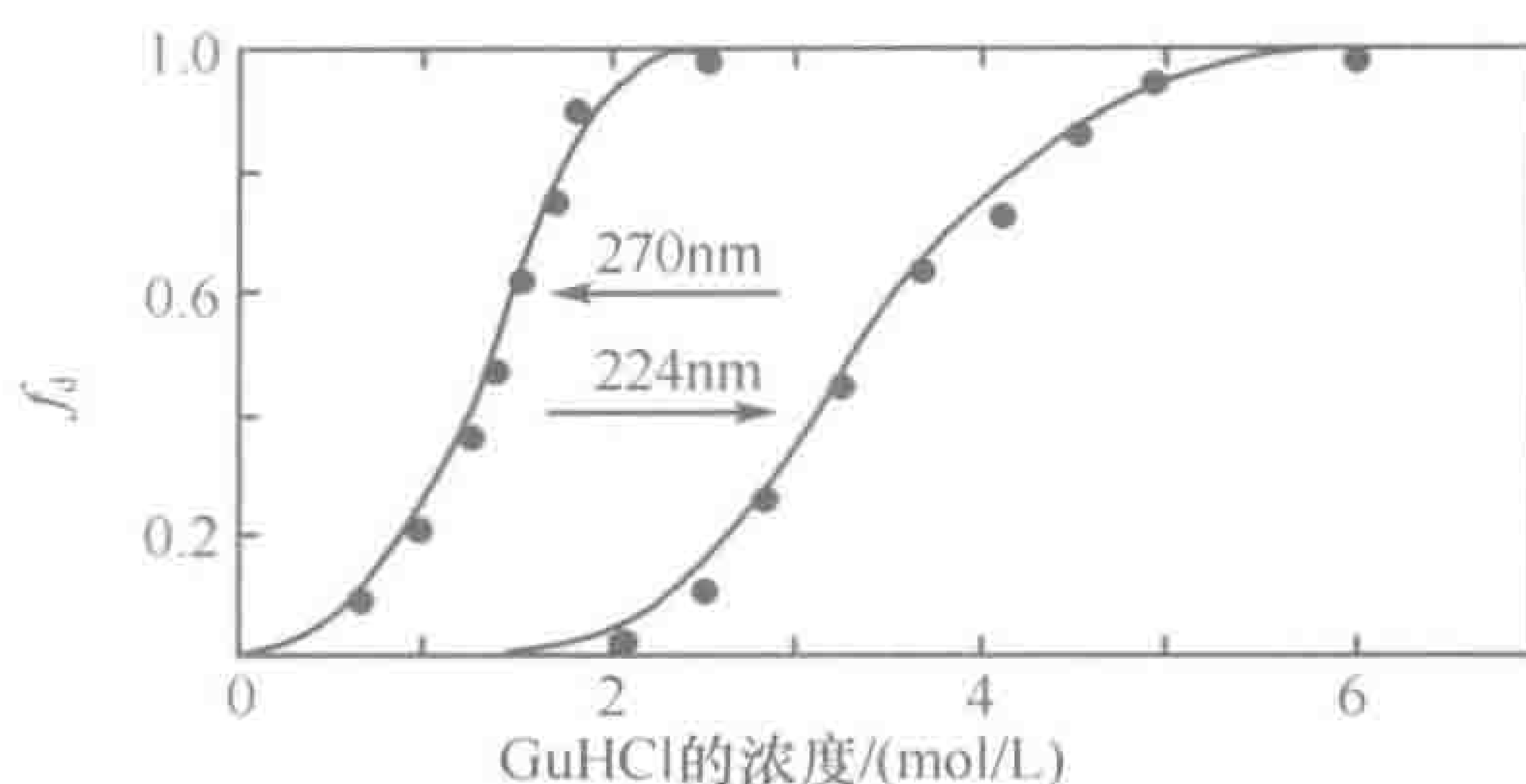


图 18-5(同图 17-8)　α-乳清蛋白随变性剂(盐酸胍)浓度改变时在近紫外区(270nm)和远紫外区(224nm)圆二色谱的相对变化。$f_d=0$ 为天然蛋白质的信号,$f_d=1$ 为完全变性分子。“中间态”存在于第一个和第二个转变之间,是(由后续研究所证明)“**熔球**”。图片取自 Pfeil W., Bychkova V. E., Ptitsyn O. B. 1986. FEBS Letters. 198:287-291,略有修改

那么,在实验上,熔球的解折叠是“全或无”型转变吗?

这里还不完全清楚。许多蛋白质的热力学研究表明,现在看来,不是“全或无”型转变。然而 Uversky 和 Oleg B. Ptitsyn 证明,熔球态有可能,至少在某些蛋白质(如碳酸酐酶①)中,被“全或无”型转变与“较疏松球”状态隔开(图 18-6)。不过,碳酸酐酶是 β 结构的蛋白质,因此所观察到的转变可能是它的 β 结构(在合成的高分子聚合物中,它能经历“全或无”型转变)。在其中观察到的用色谱法测定的转变非常缓慢,但在动力学实验中形成熔球(这将在下一讲中讨论)的过程非常快……仍然模糊不清……

在转变瞬间,“较疏松球”状态只是不如熔球那么紧密,并包含大约只有一半的二级结构,但它比线团仍然紧密得多,结构性更强得多。因此这种状态被称为“预熔球”。应该指出的是,“预熔球”和“熔球”的性质的接近性有时候导致很大的误解……

“预熔球”的进一步解折叠到线团是平缓的,正如显示在图 18-4 中的“普通”高

---

①　碳酸酐酶:1940 年发现的第一个锌酶,也是最重要的锌酶。现已报道有 80 多种锌酶,居各类金属的首位。碳酸酐酶是红细胞的主要蛋白质成分之一,在红细胞中的地位仅次于血红蛋白。含一条卷曲的蛋白质链和一个锌(II)离子。相对分子质量约为 30 000。锌离子处于变形四面体的配位环境。催化的最重要的反应是二氧化碳(碳酸酐)可逆的水合作用,使它在生理 pH 条件(pH≌7)下很快进行。上述反应对呼吸作用极为重要。

分子聚合物的"球体-线团"转变理论所需要的一样,这在前一讲中已经讨论过了。

通过蛋白质色谱学在熔球发生解体的变性浓度区域观察到了两种相("熔球"和"较疏松球")的共存。在这种色谱图上存在的分岔峰(图 18-6)表明,给定的蛋白质要么存在为一种形式(熔融的,这由其他测量手段确定),要么存在为另一种形式(某种"较疏松的"),但不会以介于两者之间的中间密度值的球体形式存在。

要证明"全或无"型转变的存在,在原则上需要蛋白质温度熔融区域的狭窄性(请回忆范特霍夫判据)。但是,范特霍夫判据并不能应用到解折叠球体的熔化:实验表明,它们远没有熔化!然而,解折叠球体的熔化通过量热法研究在一系列人造蛋白质上,甚至在生理条件下也具有熔球形式,被观察到了。对强变性剂导致的天然蛋白变性态下"熔球-线团"转变宽度的更多研究表明,许多单域蛋白质的这种宽度反比于蛋白质的大小,这更加与解折叠的协同单位是整个这种蛋白质,即转变是"全或无"型,相吻合。

■ 看来,至少在某些蛋白质中,熔球到线团的转变,开始是相变到"预熔球",而随后解折叠到线团已经是平缓地进行。但是,要解释这些转变的本质,则需要进一步的实验和理论研究。

因此,回到 Tanford 相图(图 18-7),我们看到,天然态是被"全或无"型转变(它造成这些转变相对狭窄)与其他状态(既与线团"COIL",也与温度变性状态"MOLTEN")相隔开;而且变性态必定由于有活动的侧链基团而与天然态不同,但仍可能非常紧密;最后,由温度变性态到线团的转变可能对一些蛋白质是"全或无"型的,而对另一些蛋白质则(可能)是非相变。

一点题外话。一般而言,我们不必对"不同的蛋白质的行为不同"而感到奇怪。在这里并不是蛋白质之间存在有质的差别,更在于我们是通过"实验窗口"观察各个蛋白质的,在"实验窗口"中温度范围受限于液态水的存在,而变性剂的作用范围受限于它的零浓度。因此,如含有疏水性氨基酸成分较少的蛋白质,这类蛋白质的球体很容易被变性剂破坏,在该"窗口"中只能展示到线团的转变(熔球只由剩余疏水相互作用所维持)。而那些不大稳定的蛋白质(或"被削弱"的蛋白质,如脱辅基肌红蛋白(apomioglobin),即被剥夺了血红素而稳定性削弱了的肌红蛋白),这些蛋白质显示出冷变性,总是形成线团的变性。而在更稳定的蛋白质中,观察不到冷变性,更准确些,看来是因为冷变性只能发生在水结冰的温度下,而这种温度阻止了我们的实验……

■ 现在让我们从一个稍微不同的视角来继续讨论蛋白质分子的热力学,即这样的视角:蛋白质分子需要什么样的性质,使得它不是逐渐地解体,而是以"全或无"型转变破裂?即,是什么使得它能长期忍受外部作用而不软化,然后迅速碎裂呢?(显然,这种性质有助于蛋白质工作的可靠性:如果正在工作,就正常地工作;而不工作了,则完全破裂……)

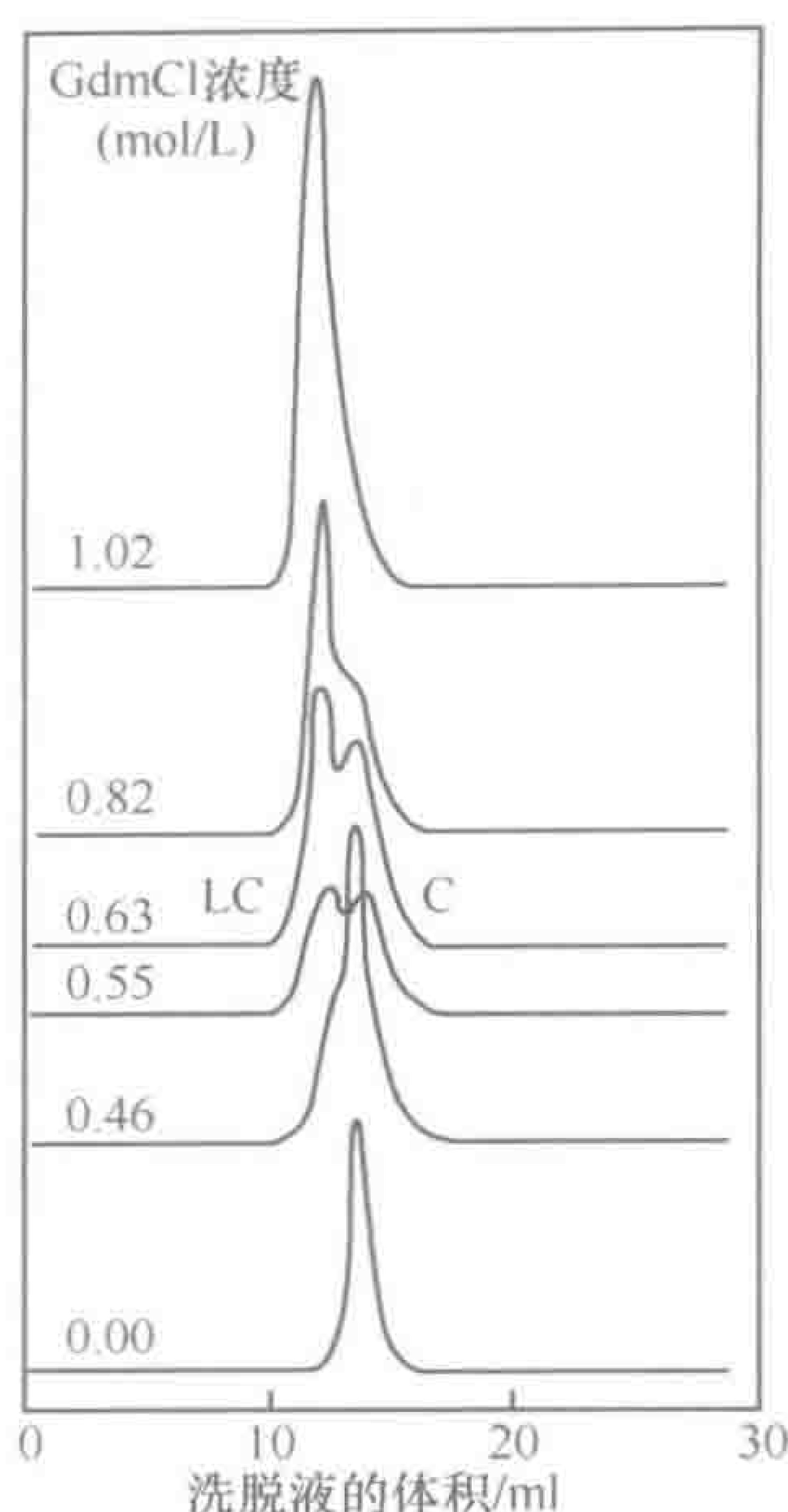


图 18-6 蛋白质色谱图上在熔球开始破裂时,变性剂(盐酸胍)浓度区域内的两相"紧密熔球"(C)和"较疏松球"(LC)共存。注意色谱图上在 GdmCl 浓度为 0.55~0.63mol/L 处的尖锋分岔。图片取自 Uversky V. N., Semisotnov G. V., Pain R. H., et al. 1992. FEBS Letters. 314:89-92,略有修改

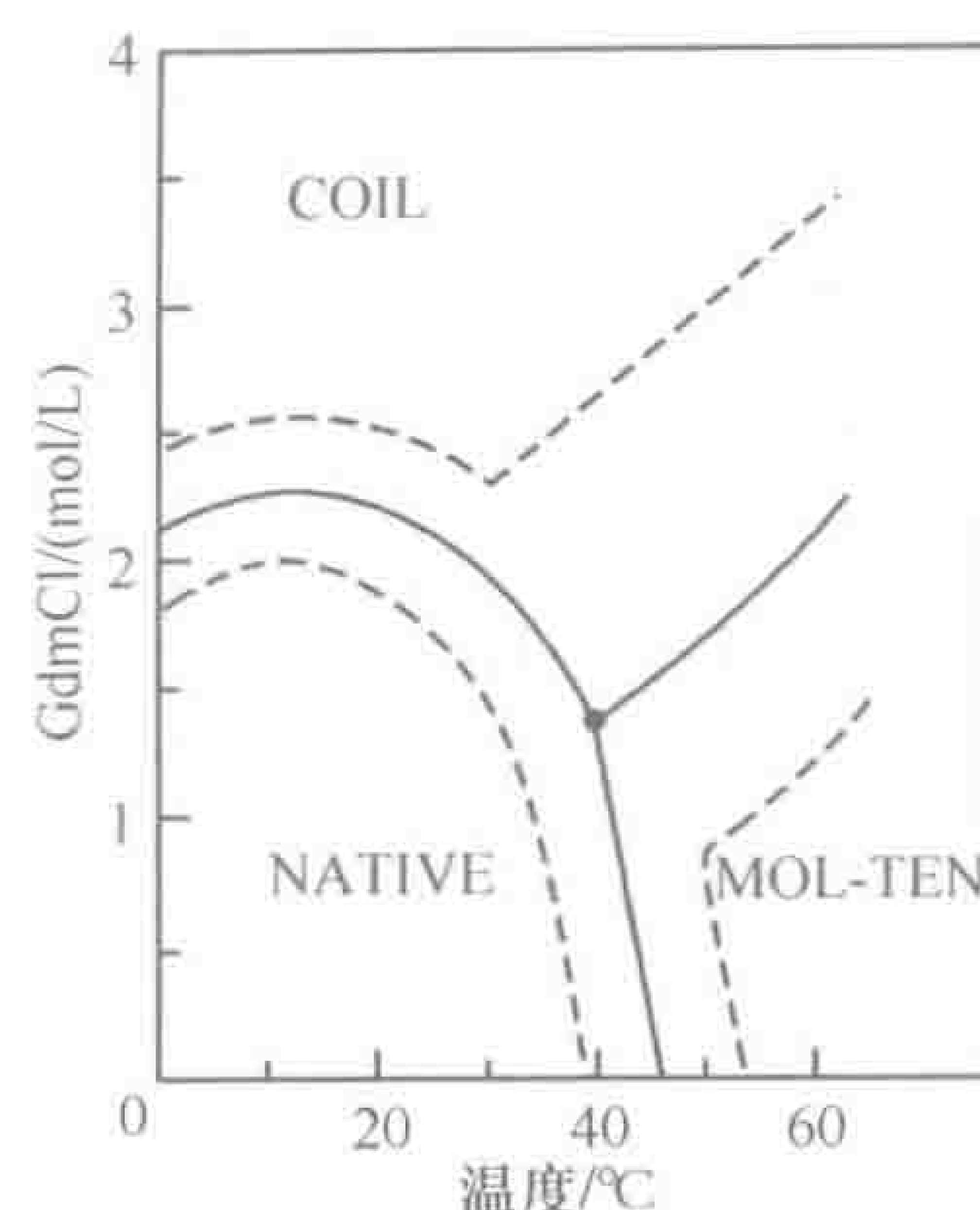


图 18-7(同图 17-6) 不同温度下溶酶体在 pH 为 1.7 的盐酸胍溶液中的构象状态的相图。NATIVE:天然状态;COIL:线团;MOLTEN:温度变性态。实线表示转变的中点,虚线为转变范围(假设从对一种状态有利的 9:1 到对另一种状态有利的 1:9;可见,COIL↔MOLTEN 的转变相当宽,与 NATIVE↔COIL 转变和 NATIVE↔MOLTEN 转变相比,"不那么协同")。图片取自 Tanford C. 1968. Adv Prot Chem. (1968) 23:121-282,略有修改

■ 利用熵 S 和能量 E 对球体密度 ρ 的关系曲线(图 18-3),我们可以建立 S 与 E 的关系曲线,甚至绘出蛋白质能量谱的特征形状[图 18-3(a)]。我提醒大家,$S(E)$ 正比于具有能量 E 的结构数量的对数值。因此,如果知道 $S(E)$,我们就可以估算在任意能量下的能量谱密度。我还提醒大家,$\frac{dS(E)}{dE} = \frac{1}{T(E)}$,其中 $T(E)$ 为与能量 E 对应的温度。

"绘出能量谱的特征形状",意味着,展示蛋白质链的不同构象按能量如何分布。在谱上我们对**什么样**的构象具有什么样的能量并不感兴趣;我们感兴趣的只是,**有多少个**构象具有那样的能量。每条谱线对应于蛋白质链的一个构象(更准确些:这里的"构象"对应于自由能的一个局部最小值,该自由能包括蛋白质内部相互作用的能量和蛋白质与周围溶剂相互作用的能量)。

所获得的蛋白质能量谱的主要特点是:最稳定的,"异常稳定的",即蛋白质的天然结构的能量与绝大多数其余蛋白质链结构的能量谱开始处的能量是分开的。正是这种分开(或者,如有些人说的,"最稳定的结构与其竞争者之间的能隙")产生了肽链的天然结构与其余结构之间的自由能垒,并使得蛋白质通过"全或无"转

变进行熔化，在此之前能“忍受”外部条件的变化而不发生改变。

能隙上方(能量更大处)实际上是“错误组装的”结构的连续谱(“连续”，是因为那里有很多线条，大约为 e^N 条，其中 N 为肽链中的链节数，而能量谱的宽度只是正比于 N)。天然结构处于能隙以下，在能隙上出现了它的几个“稍微松动的”变体。

熔化温度 T_M 对应于切线的斜率[见图 18-8(a)中的 $1/T_M$]，同时既与在能量谱最左边(“能隙后”)区域中熵 S 最小的、实际上为零的点相切，同时也与“错误组装”的结构的连续能量谱区中的高熵区相切。

除了 T_M 之外，蛋白质能量谱中还有另一个值得注意的温度：T_C。它由在处于能隙前的 $S(E)$ 曲线的斜率所确定[图 18-8(a)]，即该链“错误组装”结构的熵的收敛速度。在 T_C 温度下“错误组装”结构应该会冻结(玻璃化，因为在几乎相同的温度下这些结构数量很多)，如果蛋白质的天然结构(自动地与它们以宽能隙隔开)在甚至更高的温度 T_M 下仍然没有冻结的话。

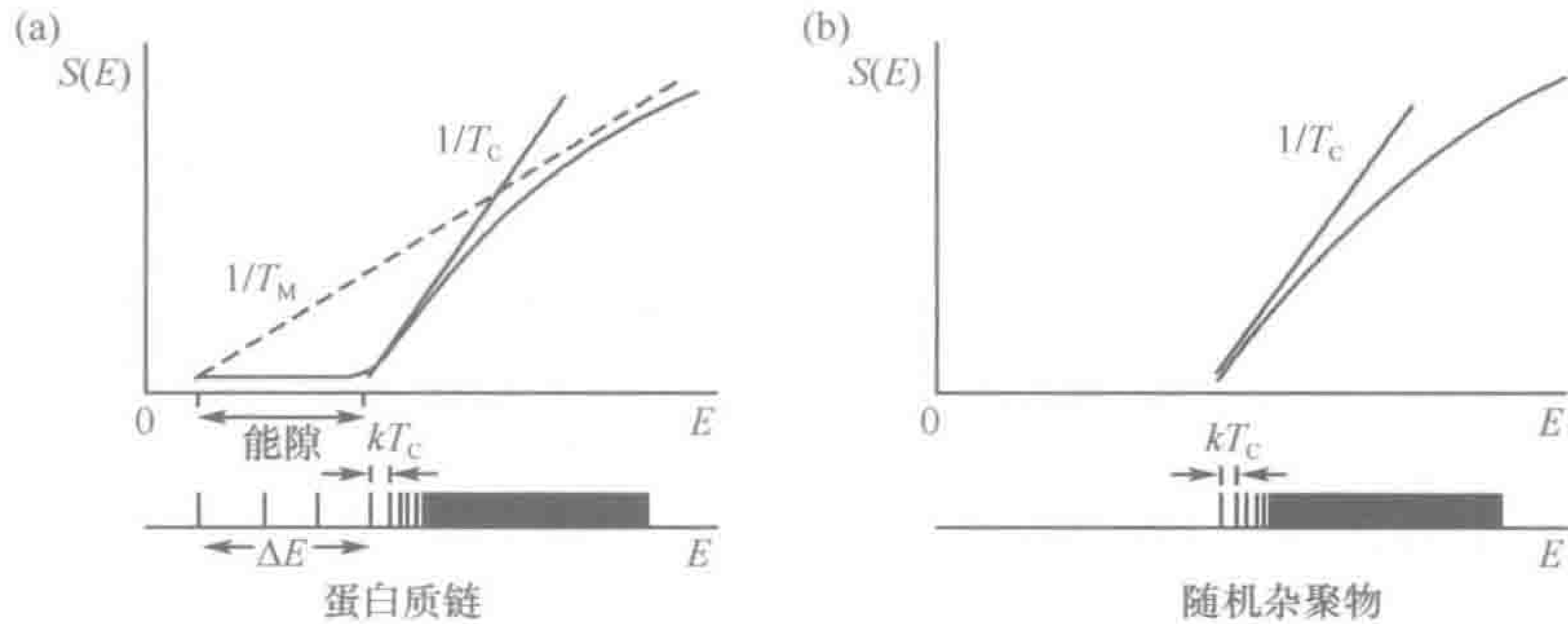


图 18-8 熵 S 与能量 E 的推测性关系曲线：蛋白质链(a)和随机高分子杂聚物(b)。图中显示了曲线 $S(E)$ 下部的切线，它确定了高分子聚合物的玻璃化温度 T_C 和蛋白质熔化温度 T_M(共存温度，即其能量最低的结构和**大量**高能量结构的自由能相等之处)。第二条切线的斜率有所减少，即 T_M 比 T_C **高很多**，即在蛋白质(“有能隙的肽链”)冷却过程中，其天然的、能量最低的结构发生凝固(跳变！)要早于位于能隙右边的“错误”结构中可能开始平缓地玻璃化！图下方显示了这两类分子的能量谱特征形状。每一条谱线(此处只显示这些线条中的一部分)对应于链的一个构象。在右侧，在这两个谱的高能部分这些线条非常多，并重叠在一起。蛋白质(“中选的”高分子杂聚物)谱的主要特征是在主要的、最稳定的肽链折叠与其他与之不相似的折叠之间存在有能隙(宽度 $\Delta E >> kT_C$)。能量落在了这个能隙中的某些结构，是链的主要折叠的稍微有点“不牢固”或“未拧紧”的变体

而“随机”高分子同聚物[图 18-8(b)]只有一个明显的温度：T_C，它的玻璃化临界温度。

异质高分子聚合物链的各种不同模型的理论物理研究表明，在随机异质高分子聚合物中没有“异常稳定”的结构，即在其中观察不到能隙[图 18-8(b)]。更准确地说，随机链中的能隙非常小，约为 kT_C，因此这些链是逐渐地冻结(在温度 T_C

下),不是像晶体或蛋白质那样跳跃性地,而是像玻璃那样逐渐地冻结。

■ 这意味着,蛋白质的相熔化必不可少的是,在肽链的最稳定结构与其竞争者之间有足够大(宽度 $\Delta E \gg kT_C$)的能隙,这种能隙并不出现在任意"随机"氨基酸序列中,而是通过挑选出"类蛋白质的"、有利于形成蛋白质的序列(这些序列只允许肽链以唯一的形式紧凑拼装成球)而产生的。

自然选择选中了这类序列,因为只有它们能够创造刚性的具有严格确定结构的蛋白质球体,而只有在这类球体中可以找到可靠地为生物体服务的蛋白质。

我们甚至设法(把描述连续谱左边区域的曲线 $S(E)$ 延长到更低能量)粗略地估算了"类蛋白质"的肽链(即具有给定大小的能隙的肽链)在所有随机肽链中的比例:

$$\text{比例}(\Delta E) \sim \exp(-\Delta E/kT_C) \tag{18.1}$$

即使能隙很明显($\Delta E \gg kT_C$),这个比率仍然非常小。这是合乎情理的:制造蛋白质很不容易!然而这个比例实在是小得不可思议:例如,在 $\Delta E \approx 10$kcal/mol 时,与 $kT_C \sim 1$kcal/mol 相比已经相当大了,这个比例是 $\sim 10^{-8}$。

现在已经借助于现代方法(如噬菌体展示技术)来查找并挑选比例这么小、但强烈而专一地与某些东西相粘连的分子。因此,对现在以氨基酸随机高分子共聚物进行的实验有时能获得"类蛋白质"的分子不必感到惊奇。不久前完成的实验表明,这样的分子占所有随机合成的蛋白质链中的 $10^{-12} \sim 10^{-10}$。这种实验[通过式(18.1)]让我们能估算出能隙值为 10~15kcal/mol。

值得注意的是,从能推导出式(18.1)的同一理论出发可以推导出,要制造同时能使两个不同的天然结构(两者是独立的,与肽链其他拼装包的能隙宽度为 ΔE)都稳定的肽链极为困难:可能性约为 $\exp(-2\Delta E/kT_C)$。换句话说,制造具有一个天然结构的肽链是奇迹,而有两个的(奇迹的平方,有三个的)奇迹的立方……因此不必奇怪,几乎没有蛋白质具有两个或更多能观察到的可替代的(alternative)天然结构(自然,这里我说的只是不同结构的共存,而不是指在蛋白质行使功能时经常能观察到的结构稍许变形,在本课程的最后几讲会讨论到)。说"几乎没有",是因为存在几种(在研究过的几千种蛋白质中)这样的"变色龙"蛋白质。在这些变色龙蛋白质中包括有:塞尔平(serpins),在没有任何聚合反应或聚集作用的情况下能非常缓慢地从活性形式转变为其非活性形式;Ltn40 蛋白,在二聚反应下它的 α 螺旋改变成部分强烈重组的 β 片;普利子蛋白(prion),我已经在第 11 讲中介绍过的一种"传染病蛋白"(虽然其替代结构的极为缓慢的形成是与其肽链转化成"感染形式"的粘联相关的);以及其他蛋白质,而且其中的一些是特殊构建的,以致具有可替代的折叠。对此我们将在后面介绍。

温度 T_C 由能隙前的 $S(E)$ 曲线斜率所确定(图 18-8);由于 T_C 依赖于大量非天然结构的性质,故 T_C 值不依赖于 ΔE 的值,而能隙,只与与非天然结构隔开的一个天然结构有关。换言之,对相同氨基酸组分的随机链和"类蛋白质"链,T_C 值是相同的。而且值得注意的是,即根据式(18.1),连续谱边缘的特征温度 T_C,恰好是在"准玻尔

兹曼"蛋白质统计学中提到的构象温度,对此我们已经在上上一讲中说到过。

图 18-8 表明,在 ΔE 并不很大时,温度 T_C **比蛋白质熔化温度略低**。而看来,很大的能隙,很大的 ΔE 值,对于分子的"类蛋白质"性质并不是必需的,因此自然选择未必强求太大的 ΔE 值。而同时,根据式(18.1),大 ΔE 值的可能性很小。不管怎样,这又一次有待于实验结果……

■ 能隙是蛋白质物理中重要而基础的东西。它不仅对于蛋白质只能以相变形式发生破裂,即它能忍受(到已知的极限)外部条件的变化而不发生改变,即能保证蛋白质工作的可靠性是必不可少的,而且我们很快还将看到,它对于蛋白质的天然结构能快速而正确无误地折叠也是必不可少的。因此在实验上对这些能隙进行估算(只可惜,这还没有做)将会是非常有趣的。

第19讲

蛋白质结构在体内(*in vivo*)和在体外(*in vitro*)的形成。在体内自动折叠中的辅助机制:共翻译折叠,分子伴侣及其他。自发的自动折叠在体外有可能。“Levinthal 佯谬”。非细胞系统中的蛋白质折叠实验,以及关于单词“在体外(*in vitro*)”的不同理解。蛋白质自动折叠的分级机制。许多蛋白质中发现了亚稳的(积聚的)折叠中间体。熔球——通常(但并非必须)在天然条件下观察到的蛋白质折叠中间体。某些蛋白质的最简单(单步骤)折叠,没有任何积聚的亚稳中间体。膜蛋白的自动折叠。

今天,我们将来讨论蛋白质链及时地获得其天然结构的过程是如何进行的,并讨论在蛋白质自动折叠过程中出现的中间体。我将只讲及水溶性球状蛋白。对膜蛋白的自组织的研究相对比较少,我对此将只简单提及。纤维蛋白的自动折叠研究得多一些,但不是在物理层面,而是在生物化学层面(对此已经在胶原蛋白如何形成的例子中讨论过了)。

■ 在活细胞中,蛋白质在核糖体上合成。蛋白质链生物化学合成约需几分钟,而“齐备的”已折叠好的蛋白质的整个制作占用时间也大致是这么多,实验上未观察到两者之间存在差别(参考:细菌的整个生命周期可能持续只有几十分钟)。因此,我们可以自然地认为,蛋白质的折叠可能甚至在核糖体上整个蛋白质链还没有完全合成就开始了。

遗憾的是,在体内(*in vivo*),即细胞中,关于蛋白质的形成的可靠的实验数据非常贫乏:在整个细胞质背景下看到新生蛋白质的结构变化非常困难。通常必须停止合成,分离出“尚未做好”的蛋白质,单独进行研究:整个需要花费几十分钟,在此期间所研究的蛋白质的空间结构可能会发生重大改变……

以大蛋白质进行的一系列这类实验表明,它们的第一个域,即N端的域,来得及在肽链完整合成结束之前折叠(或更确切地说:实验本身所能看到的长度,具有折叠的能力)。

然而,在此必须强调,所有这些数据是关于**多**域蛋白的,这是一项重要的事先声明,因为“自动折叠单元”看来,不是蛋白质整体,而是它的独立域,即没有观察到“半折叠的”域,我们也不能说,首先是它的N端那一半开始折叠,然后再是C端那一半。

折叠单位是域,这是由两组数据证实的,虽然不是在体内、而是在体外(*in vitro*)获得的:①单独取出的域常常能够正确地自动折叠;②在C端去掉了大约十

来个氨基酸的单域蛋白质不能够自动折叠。在另一方面，A. A. Komar 及其同事证明，在非细胞系统中合成的血球蛋白链，只要核糖体制作成该链的一半稍多一点，就能够与血红素相结合。[根据其结构和其在体外的熔化特征(即，它的变性是按“全或无”进行的)来判断，血球蛋白是一种单域蛋白。不过其 N 端那一半能形成紧密的亚域，在该亚域中有与血红素相结合的肽链部分。]

从**非细胞**蛋白质合成系统中(即不完全是在体内，也不完全是在体外)获得了一系列关于蛋白质折叠的有趣的数据。这种系统是由核糖体、转运 RNA 和信使 RNA 以及对于蛋白质链基质合成必需的其他辅因子所构成的。

这里有必要提请注意，说到在体内(*in vivo*)和在体外(*in vitro*)上的实验，物理学家和生物学家常常有不同的理解：因为在纯粹的体内(在生物体中)和纯粹的体外(“在玻璃容器中”)之间存在有很多允许双重解释的级别。例如，对于非细胞系统(含有全部核糖体、起始子和终止子、分子伴侣等)，从物理学家(对他们来说，“体外”是溶剂中的单个蛋白质；而非细胞系统……已经有这么多的生命反应！)的观点来看显然是在体内的实验，而从生物学家(对他们来说，“在体内”是在有生命的，最好是完整的生物体中)的观点来看这是在体外的实验。然而关于生物分子结构的实验，如在生物体中对单个蛋白质的研究，实际上是不可能的。因此在实践上总是有合理的折中：对于在体内的有生物学意义的事件通过用实验上可进行观察的在体外的事件来近似。

遗憾的是，在非细胞系统中也很难看到在巨大的核糖体背景上看到新生蛋白质的**构造**。确实，**构造**是很难看到的。但是，如果幸运的话，可以看到刚刚合成的蛋白质的**活性**。

令人高兴的是，这里可以使用荧光素酶：这种蛋白质在采纳其天然构象之后，能催化带发光的反应①。因此，荧光素酶的出现是很容易观察到的，因为在细胞中没有其他东西能发光。通过研究这种蛋白质的生物合成和折叠，在俄罗斯科学院蛋白质研究所里的 V. A. Kolb，E. V. Makeev 和 A. C. Spirin 证明，第一批活性蛋白出现在启动其生物合成反应的 10min 之后，而反应被关闭或阻断之后马上停止生成活性蛋白(图 19-1)。这意味着，**已经**合成、但**尚未**折叠的荧光素酶链几乎是不存在的，即我们看到，这种链中包含超过 500 个残基的大蛋白质的折叠要么是与翻译同时进行(共翻译，cotranslation)，要么是在合成之后几乎瞬间进行的。

的确，最近完成的实验(Eichmann，et al. 2010)表明，由核糖体合成的蛋白质链不到结束是不会形成既不挂在核糖体上也不从它上面脱落的确定空间结构的(原则上与该链形成熔球是不矛盾的)。于是，在体内由核糖体合成的蛋白质链在结束之前表现得像在体外截短了几个 C 端氨基酸的链：它并未形成确定的三维结构。

---

① 萤光素和萤光素酶是对所有能够产生萤光的底物和其对应的酶的统称。萤光素与氧气反应时会发光。在没有萤光素酶的情况下，这种反应的速率非常慢。萤光素酶能催化这种反应。

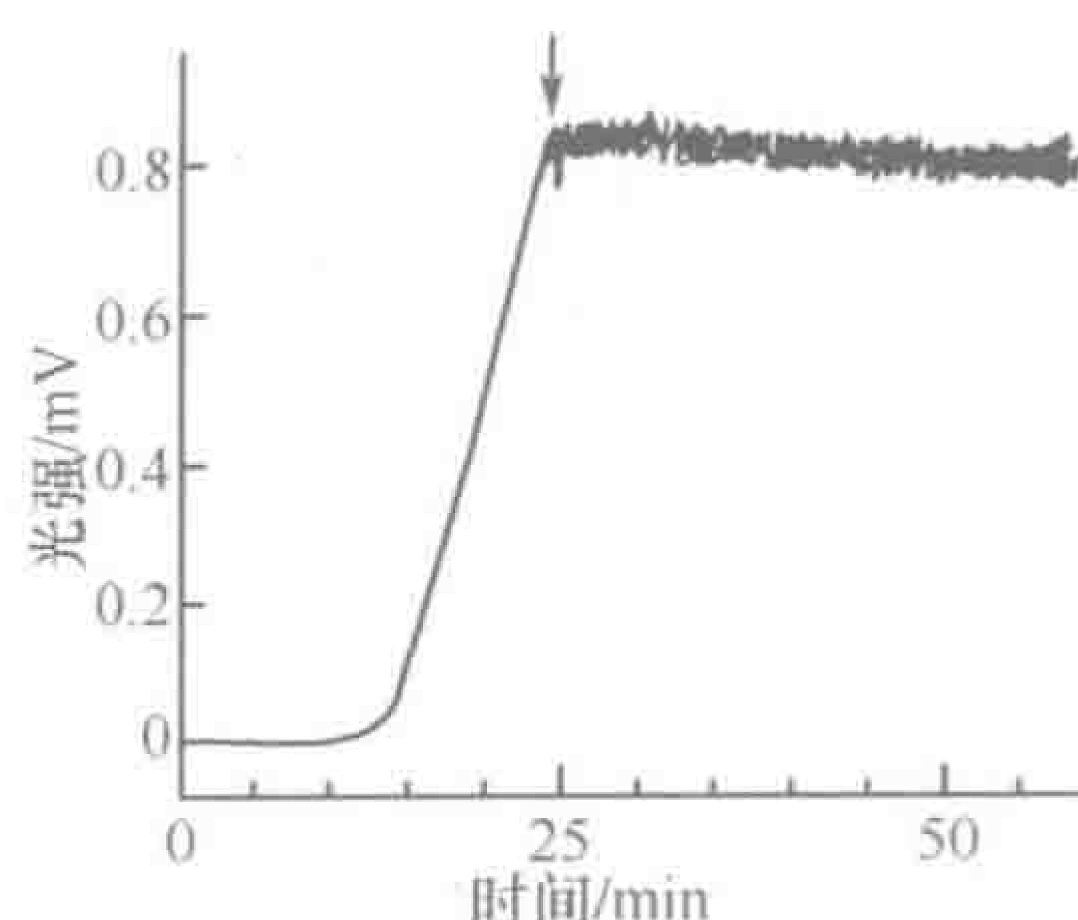


图 19-1　在非细胞体系中合成的天然荧光素酶的发光。“0”时刻荧光素酶合成反应启动。反应关闭时间如箭头所示。在此之后发光停止增强,即未观察到产出新的活性蛋白质。图片取自 Kolb V. A. ,Makeev E. V. ,Spirin. A. S. 1994. EMBO J. 13:3631-3637,略有改动

■　然而,为了不仅说明蛋白质自动折叠结果,还要阐明折叠的过程,必须研究蛋白质“完全**在体外**”(无任何核糖体、伴侣等)时的自动折叠,即蛋白质在溶剂中的复性。

大约在 20 世纪 60 年代有一个惊人的发现:如果球状蛋白在生物合成之后没有受到强烈的化学修饰,**在体外**就能够自发地自动折叠(复性)。在这种情况下,被温度和溶剂等造成“轻微”(链未断裂)破坏的蛋白质结构能自发地在“正常的”环境中恢复。不过,有效的复性需要精心选择实验条件,否则,它会被蛋白质的聚集所阻止。

课堂里的声音:会不会是在变性蛋白质中保存了从空间结构的天然状态中遗留得的某些性质,能指导它的折叠沿正确的路径进行?

讲课人:如果蛋白质是被“按照要求”进行变性的,则不可能。当然,如果你在变性蛋白质中不保留它的完整的二硫键,而是破坏它们;而且你等待几分钟,以足够消除对顺式(*cis*)和反式(*trans*)脯氨酸(Pro)天然形式的记忆。而且你把蛋白质放入到足够强烈的变性剂,使得它变成线团(二级结构被破坏,体积成数量级增大,即所有沿链的远程接触都消失了)。

课堂里的声音:然而,有相当可靠的数据,说明即使在完全解折叠的蛋白质中仍然存在有相当数量的非沿链相邻的残基之间的“天然”接触!

讲课人:有这样的数据。不过(暂且不谈含有二硫键的蛋白质),它们涉及只是在肽链上相当接近、仅由几个残基隔开的残基之间的接触。而主要的问题(就我们现在所讨论的而言)在于:它们是从蛋白质的天然结构中遗留下来并被冻结了的,还是它们仅仅是在解折叠链中时不时地自发形成的呢?遗憾的是,核磁共振(NMR)(而基本上正是它看到了所讨论的接触)不能区分“冻结的”接触和时不时地形成的接触。原因在于,核磁共振的信号随 R^{-6} 衰减(其中 R 为接触距离),因此同样的信号对应着距离 R=4.5Å、存在时间 100% 的接触,也对应着距离 R=3.0Å、存在时间 10% 的接触。我个人认为,所观察到的接触不是冻结的(当然,有冻结的非共价二硫键;但我们在这里不讲及它)。我看不到对此的任何物理原因。我认为,所讨论的接触仅仅是在互相吸引的链块之间经常形成的(记得吗,基本上也正是这种互相吸引的链块在天然蛋白质中发生接触)。显然,这也同样适用于偶尔在圆二色性(CD)谱中看到的解折叠链的“残余”二级结构。

■ 蛋白质自发自动折叠的现象,是 Anfinsen 研究组于 1961 年发现的。这是以牛核糖核酸酶的生化活性和“正确的”二硫键自发恢复为例而发现的:把它完全解折叠并破坏所有二硫键,然后放到“天然”条件下。这一发现随后在许多其他蛋白质上得以证实,还由纯化学合成的蛋白质链能自发进一步折叠为活性蛋白质(Merrifield 等的实验)得以证实。这一发现在一级近似上把蛋白质的结构形成研究与蛋白质链的合成研究区分开来。

■ 蛋白质在体外的自动折叠可以视为最简单的(因此,对于我作为物理学家而言,是最有意思的)“纯”**自组织**(self-organize)的现象,因为没有谁来帮助蛋白质。

一般来说,蛋白质(以及 RNA)空间结构的自动折叠是唯一的在“非生命”的自然界中没有类似物的物理现象。这种自组织行为让人想到晶体的形成,但这种晶体,首先,不具有周期性空间结构,其次,构造极为复杂,最后,非常小。

从物理学的观点来看,蛋白质结构的自动折叠属于“从有序中产生有序”的现象(按照普里高津①的分类):蛋白质结构的三维“非周期性晶体”(薛定鄂语)早就根据其链中的氨基酸固定顺序而生成了。显然,应该立即指出,蛋白质(和 RNA,以及一般的晶体)的三维结构的自组织②产生于分子对热力学平衡的趋向,因而不同于“从无序中产生有序”型的自组织(无论是在 Belousov-Zhabotinsky 的振荡化学反应中的自组织,或是“捕食者-猎物”生态系统中的自组织),后者通常是在能量流上工作的非平衡系统中的自组织。

■ 然而,在转向讲述蛋白质在体外的自动折叠、讲述这种自发过程的物理学之前,我简短地提及细胞用以提高蛋白质自动折叠效率的机制。

核糖体逐渐地、不完全均匀地制造并释放出蛋白质链:有一些暂停,肽链的生物合成在“罕见的”[在细胞的信使 RNA(mRNA)上也相应地“罕见”]编码上会短暂停顿。有人认为,暂停对应于域结构的边缘,有助于域结构不受外来干扰、安静地成熟。

此外,蛋白质链在细胞中折叠时是在特殊蛋白质(分子伴侣)的照顾下进行的。分子伴侣能防止蛋白质聚集。其中许多是由细胞对热休克的响应下而生产的,因为在温度升高时能引起聚集的疏水力会增大。

如果蛋白质在细胞中聚集,它们常常形成“包含体”。蛋白质在包含体中不具有天然结构(尤其是,它们的二硫键无序地结合形成)。然而,如果拆散包含体,从中可以得到天然蛋白,通过体外复性。

在细胞中分子伴侣与能导致聚集(在浓稠的“细胞汤”中是多么自然)的麻烦

---

① 伊利亚·普里高津(Ilya Prigogine,1917—2003),比利时化学家、物理学家,非平衡态统计物理与耗散结构理论奠基人。由于他在“非平衡热力学,尤其是在他的耗散结构理论”方面的成就,被授予 1977 年诺贝尔化学奖。

② 本书俄文版中把蛋白质自动折叠都称为“自组织”,为方便理解,翻译时大部分都译为“自动折叠”。

进行斗争。“小”蛋白伴侣,如 hsp70[热休克蛋白(heat shock protein),分子质量为70kDa],能与蛋白质相结合,保护它不聚集,然后松开它(消耗 ATP)。“大”分子伴侣,如 GroEL/GroES 或 TriC 的工作对象主要是多域蛋白,而且尤其是域由远程链块组成的那些蛋白质。这些分子伴侣形成由 GroEL 蛋白制成的试管(直径约几个纳米,还有由 GroES 蛋白做成的盖子,时而与 GroEL 黏附,时而与之脱落)。展开的(或解折叠球状态的)蛋白质链就位于(在体外)在这些试管中(或者,更快一点,在试管入口,如 G. V. Semisotnova 的实验所表明的)。看来,它们在体内就变成新合成(在遇上 GroEL 之前被分子伴侣 hsp70 和/或 hsp40 所黏结)的蛋白质或蛋白质域。根据当今流传的观点,这种“试管”(有时被称为“Anfinsen 笼”)保护新生的蛋白质不参与聚集,也不受浓稠的包含其蛋白质酶的细胞“汤”的作用等。而且,有人认为,这种试管时不时地主动“抖动”(这时在 GroEL 中发生构象转变,增大或减小其疏水表面),即它打开(消耗 ATP)和关闭,但只有在蛋白质已经折叠并不再黏结到“试管”上时,它才释放蛋白质。然而,G. V. Semisotnova 研究组用工作证明,其实 GroEL 仅仅只是从溶液中收集“多余的”无结构的蛋白质,使得少量的剩余在溶液中的蛋白质避免聚集,并在与 GroEL 蛋白伴侣没有直接相互作用时正确地折叠。因此,无论怎么样,蛋白伴侣只是作为蛋白质折叠的“孵化器”而工作,不过,如最近 A. S. Spirin 研究组所示,它们有时也能作为“冷藏箱”那样工作,能延缓蛋白质的折叠,直至把蛋白质拼装成四级结构或转运到细胞的适当地方为止。

此外,蛋白质的自动折叠可以被几种酶加速,如脯氨酰肽异构酶(主要是催化自身缓慢进行的脯氨酸反式-顺式转变;有时候正是这种转变限制了折叠速度),或二硫键异构酶(催化二硫键的交联和断裂)。

■ 在活细胞中,蛋白质的折叠是在拥挤的大分子环境中进行的。

因此,根据理由充分的 C. Dobson 的理论,折叠过程的关键步骤在于熔球:蛋白质链在生物合成过程中形成的就是它,受分子伴侣保护避免聚集的就是它(如果没有保护好,正是由它形成淀粉样蛋白和其他常常是致病的聚集体),天然蛋白质也正是由它所形成的。

在进化过程中产生了一系列机制,以避免由蛋白质在细胞中的拥挤环境带来的问题。这些机制包括各种分子伴侣家族的作用,以及共翻译(cotranslation)折叠。与翻译后折叠相比,共翻译折叠的重要性大为增加,尤其是在主要是包含多域蛋白质的真核细胞中。然而尽管如此,看来,不存在任何理由可以认为,在氨基酸序列之外还存在什么东西来维持蛋白质(在任何情况下,无论是单域或多域)在细胞环境中的可靠构象,尽管还有一些研究人员的持有相反的看法。

从不久前的研究工作中得出的来一个值得注意的想法,是随着真核细胞由原核细胞组合而出现,以及更大更复杂的蛋白质(典型的真核蛋白质由 4~5 个域组成,而原核蛋白质由 2 个域组成)的产生,而发生了从翻译后折叠到共翻译折叠的“进化飘移”的假说。

共翻译折叠，即新生的蛋白质在尚未脱离核糖体时的折叠。由于最近核糖体结构研究在原子分辨率级的重要成果，共翻译折叠引起了人们极大的兴趣。新生多肽的折叠可以由“隧道”核糖体所易化，有可能该肽最初处于“隧道”核糖体内，受到保护避免聚集或降解。此外，有一些（的确，还不足够）数据表明，核糖体本身和核糖体的某些部件，特别是其大亚基，尤其是其 23S 的 RNA 可以作为分子伴侣起作用，可以促进和加速事先已变性的蛋白质的折叠。核糖体的这种作用可以额外（除了对位于其中的蛋白质仅仅提供免遭聚集的保护之外）为下述事实提供一种解释：在体内的蛋白质折叠是它们在核糖体上进行生物合成的同时进行，与之相比，在体外（无核糖体）的变性大蛋白质的折叠通常进行得慢得多。

然而，在体外的蛋白质折叠实验表明，所有这些细胞中的分子伴侣的工作可以通过选择相应的外部条件（低浓度的蛋白质，合适的氧化还原电势）来替代。这种替代不改变折叠结果：如果蛋白质不掉落到沉淀物中，而是在体外折叠，则它折叠成的结构与在体内折叠是一样的。确实，这常常需要比在体内的自动折叠花费更多（而有时更少）的时间，但结果是相同的。

不仅如此：已经知道，小蛋白质链可以完全在体外（试管中）化学合成（除了从 C 端而不是从 N 端开始合成），则它能折叠成正确的空间结构，产出率很好。

这一切都证明，对于建造蛋白质三维结构的所有必不可少的信息都包含在其肽链中的化学氨基酸序列之中。

看来，随了蛋白质链合成本身之外，细胞的生物合成设备（核糖体+分子伴侣+…）对于蛋白质空间结构的折叠只是一个孵化器似的：这个“孵化器”并不决定蛋白质的结构，但是为它的成熟提供条件。就像普通的孵化器保障禽蛋孵化，但不能事先决定孵化出来的是小鸡还是小鸭。

课堂里的声音：所有这些，严格来说，主要适用于小型水溶性球状蛋白（和大蛋白质的域）。对大蛋白质，特别是高等生物的蛋白质，要更复杂一些：它们远非全部都能自发复性。至于膜蛋白和纤维蛋白的自发自动折叠，只是在一些蛋白质上发生，但通常它们不能达到完全的复性……

讲课人：说到那些“难”复性的蛋白质：有可能是问题出在蛋白质聚集上（如某些膜蛋白尽管在水中不能复性，但能在变性剂中复性）；也有可能在于翻译后修饰，特别是在真核细胞中。我们约定，在这里我们只考虑相对不大的球状蛋白。让我们先开始来讨论它们的自动折叠……

■ 蛋白质（和 RNA）的自发自动折叠现象的神秘性被总结为“Levinthal 佯谬①”。说明如下。蛋白质链有极多的可能构象（每个氨基酸残基有约 10 个可能的构

---

① Cyrus Levinthal（赛勒斯·列维萨，1922—1990），美国分子生物学家，化学信息学创始人，1969 年提出此佯谬。

象，则由 100 个残基构成的链有约 10^{100}个可能构象）。因此，蛋白质应该在这大约 10^{100}个可能的空间构象中寻找“自己的”那一个。蛋白质只有直接达到各个构象，才可以“感觉”构象的稳定性，因为甚至 1Å 的偏差都可能强烈地增大紧密蛋白质球中的肽链能量。因此如果从一个构象转变到另一个至少花费 ~10^{-13}秒的时间，则遍历全部 10^{100}个构象应该需要花费约 10^{80}年，而我们宇宙的存在时间也只有 10^{10}年，无穷小的数值……问题是：蛋白质怎么能在几分钟内“找到”自己的结构？

佯谬归结如下。一方面，按照所有的测试，天然结构表现为在所有肽链结构中最稳定的：蛋白质链经过不同的动力学过程[在生物合成过程中在核糖体上折叠，在转移穿过细胞膜之后折叠，在试管中折叠（复性），无论它是否是在该试管中解折叠的（或者仅仅是由化学方法合成的）]都达到了天然结构。而另一方面，不可能有任何办法保证这种结构是在所有可能结构中最稳定的：因为正如 Levinthal 的估算所说明的，蛋白质链就根本没有时间对此进行确认！

“蛋白质怎么在无穷多的可能结构中挑选出它的天然结构？”Levinthal 这样问，并回答说：“看来，蛋白质的自动折叠是沿着某个特定的‘折叠路径’进行的，而该路径结束时的结构就是它的天然结构。”换句话说，Levinthal 认为，蛋白质的天然结构不是由稳定性决定的，不是由热力学决定的，而是由动力学决定的，即它对应的不是全局的肽链自由能最小值，而只是快速达到的最小值。

究竟是什么——动力学还是热力学决定蛋白质链的折叠，绝对不是一个纯思辨的问题。它在解决蛋白质物理学具体问题的过程中经常出现，无论是根据蛋白质氨基酸序列来预测其结构（需要知道预测的是什么：最稳定的结构还是折叠最快的结构？）或是设计新的、在自然界不存在的蛋白质（需要知道在做什么：最大地增强所期望的结构的稳定性，还是铺设最快的折叠路径？）。

在解析出第一个蛋白质三维结构并发现自动折叠现象之后，很快就开始了对蛋白质折叠机制的讨论。首先出现的大概是 Philips 假说，它认为在新生蛋白质的 N 端（蛋白质生物合成是从该处开始的）出现了结构的胚芽，肽链的其他部分缠绕在它上面。直到现在这种观点还以这样或那样的形式出现在一些研究工作中。然而这一假说没有得到证实。Goldenberg 和 Creighton 的出色工作表明，在体外的自动折叠中，链的 N 端并不起决定作用，至少对于单域蛋白质是如此。实验表明，小蛋白质（胰腺蛋白酶抑制子）的闭合成环的链仍然具有自动折叠的能力；甚至剪断该环，使原肽链中间形成新的 N 端，自动折叠仍能达到原先的空间结构。在当今，制作重新排列成环状的蛋白质已经是很平常的了。

■ 在寻找解决蛋白质自动折叠问题的努力中，O. B. Ptitsyn 于 1973 年提出了蛋白质分阶段折叠的方案（图 19-2）。这一假说后来称为“构架模型”（framework model）。在世界上的科学文献中，对蛋白质自动折叠问题很大程度上是在这一假

说的框架内进行了更深入的讨论①。

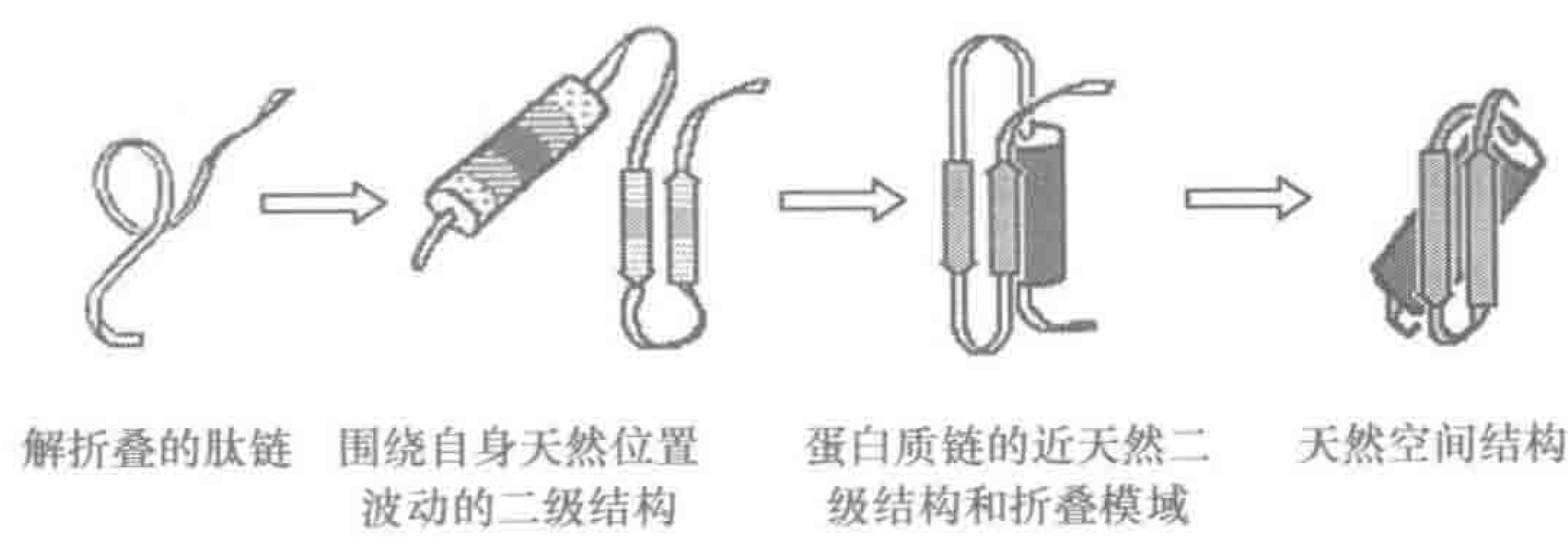


图 19-2 O. B. Ptitsyn 提出的(1973 年)蛋白质折叠的分阶段模型。图中标出了二级结构:α 螺旋(圆筒)和 β 片(箭头)。预言的两个中间体后来在实验中找到了,并被称为“预熔球”和“熔球”

该模型假设,在蛋白质结构的形成中,不同的相互作用是按顺序参与的,并强调在不同折叠阶段中形成 α 螺旋和 β 片晶核的重要性,强调这些晶核粘合为大致类似天然球的球状,以及在自动折叠的最后步骤中球体构造的最终“修整”。

这一方案的最重要的基础在于,虽然纯属假设、但现在已经进入教科书、现今以“熔球”而闻名的蛋白质链状态。

真正“从笔尖”上产生的熔融蛋白质球后来被 Dolgikh, Semisotnov, Gilmanshin, Bychkova 等于 20 世纪 80 年代在 O. B. Ptitsyn 实验室中在实验上观察到了,它最初是作为“轻微变性”蛋白质(对此我在前一课中讲过)的平衡状态,然后是以在体外的蛋白质折叠的动力学中间体,并进行了详细研究。

例如,在碳酸酐酶 B 的复性过程中发生了熔融蛋白质球的积累(图 19-3)。在这种实验的过程中,变性蛋白质的初始线团状态通过在高浓度变性剂的溶液中人工培育而得到,而复性是通过把该溶剂在大量水中快速稀释而实现,即所观察的复性过程是从线团开始并结束于天然蛋白质的恢复。然而,所复性的蛋白质的各种性质以不同的速度发生变化,这证实了蛋白质分子的某种“中间”态的积累,在复性过程开始处。

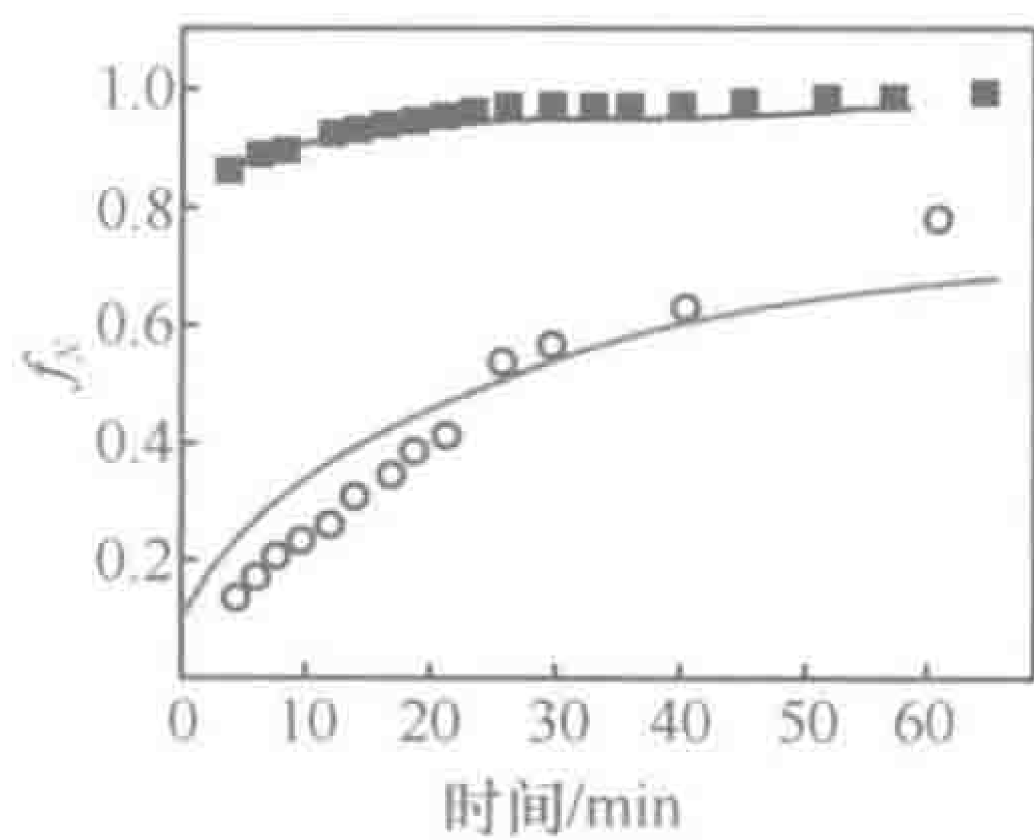


图 19-3 碳酸酐酶 B 复性过程中的“天然度”f_N 的恢复动力学。蛋白质从完全解折叠状态(存在于 5.45mol/L 盐酸胍溶液)到天然状态(0.97mol/L 盐酸胍溶液中)的转变以不同参数进行观察:内部黏度(■■■),222nm(——)和 270nm(——)的圆二色性,以及酶活性(○○○)。图片取自 Dolgikh D. A., Kolomiets A. P., Bolotina I. A., et al. 1984. FEBS Letters. 164: 88-92,有简化

可见,在复性过程中,内部黏度(表征球体的体积)的“天然”值和 222nm 下的圆二色性(表征二级结构的存在)的恢复(更准确些:**几乎**恢复)比 270nm 下的圆二

① O. B. Ptitsyn 是本书第一版的第一作者,这句略带谦虚的话的意思是,世界科学界接受了他的这一假说。

色性(表征蛋白质三级结构中侧链基团的拼装)或酶活性(是其二级结构的天然性的量度)要快得多。这说明存在有动力学中间体的大量(即亚稳的——"类似稳定的":因为它在明显的时间段内存在)积累。而且这种中间体是解折叠的球体:它的紧密度和二级结构接近于天然蛋白质,然而它不具有任何侧链基团的天然拼装状态(根据"近"紫外的 270nm 的圆二色性来判断),也不具有天然("固态的")蛋白质的酶活性。

熔球,是(强调一下:在近似生理条件下)许多球状蛋白在体外从线团开始的自动折叠的起始路径中的早期过渡态。这样的中间体只要几毫秒就能形成,而由 100~300 个氨基酸残基构成的天然蛋白质的性质要完全恢复则需要几秒钟(对某些蛋白质)至几小时(其他蛋白质)

必须强调的是,最慢的(限制速度的)自动折叠步骤不是发生于早期阶段,不是在于形成熔球,而是由熔球形成紧凑拼装的天然球体(图 19-3)。

■ 熔球并不是在蛋白质折叠中观察到的唯一的中间体。在对发生在折叠的第一毫秒期间(即在形成熔球之前)的现象能够进行研究之后,还观察到带有部分形成二级结构的状态和带有部分浓缩的链(即上面提到过的预熔球)的状态。在含二硫键的蛋白质中测定了多种中间体,它们能帮助跟踪二硫键的形成过程,如此等等。

应该说,蛋白质折叠的"动力学控制"假说极大地促进了对折叠中间体的研究。这一工作方向的主要思路在于,若能捕捉到中间体,我们就能够追踪蛋白质的折叠路径。在化学和生物化学中这样通常能帮助人们理解复杂的反应是如何进行的。因此,正如人们所说的,研究工作沿着"化学逻辑"的轨道进行,它的迫切命令是:"寻找并挑出中间态!"

但是在蛋白质自动折叠的研究中,这种逻辑只是部分地起作用:对极多的蛋白质都已经确定了中间体(熔球),但关键问题是:蛋白质链如何能够在非常短的时间内找到自身的天然结构 ? 这仍然没有回答。

■ 通过研究小蛋白(由 50~100 个残基构成),已经取得了一些认识上的进展。看来,其中一些在体外折叠时没有观察到在实验中发生"积累"的中间体。而且不仅在于特别创造的接近于蛋白质熔融点的条件下(已知变性和复性热力学对应着"全或无"型转变,即缺乏任何明显数量的"半折叠"形式的蛋白质,由此出发,在这种条件下可以先验地(*a priori*)期待有单步骤折叠动力学)未观察到中间体,而且(对这些小蛋白)在"生理"条件下也未观察到:在生理条件下大部分蛋白质的折叠都经过熔球型中间体而进行(图 19-4)。

对既无二硫键也无顺式脯氨酸(即先前广泛用于寻找折叠中间体的那些特性)的小蛋白的研究表明,这些被剥夺了"不必要的复杂性"(二硫键和顺式脯氨酸)的蛋白质有时候折叠得特别快,特别是在纯水中(或在生理溶剂中),不仅远快

于大蛋白(这并不令人惊奇),而且并不比那些折叠过程中存在有中间体(看来应该能加速并易化折叠)的小蛋白要慢……

"单域"尺寸的蛋白质在生理溶剂中折叠(能观察到积累的中间体)通常花费几秒或数十秒,但某些这类在折叠时没有中间体积累的小蛋白能在数毫秒内甚至在不到一毫秒内(确实,从最小的蛋白质到长达100个残基的都是如此)达到天然结构。这些蛋白质的折叠过程中缺少中间体,是根据蛋白质中二级结构、侧链基团的拼装、色氨酸荧光(依赖于色氨酸埋入到球内的程度)和氢交换率等性质的恢复速度都完全相同而得到确认的。

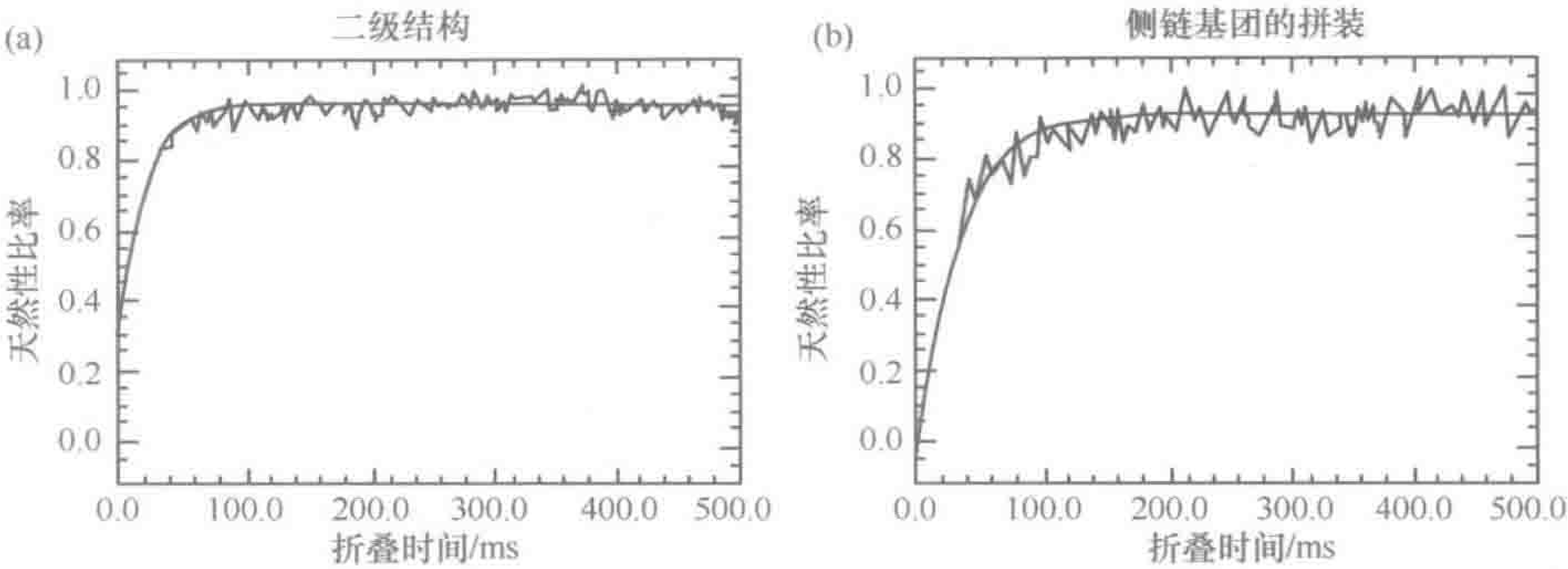


图 19-4 ACBP(乙酰辅酶A结合蛋白)的单步骤复性。记录了225nm(a)和286nm(b)的圆二色性的恢复。前者反映了二级结构的恢复,后者反映了三级结构(更准确些:侧链基团的拼装)的恢复。细折线为实验信号,粗线条为考虑到实验误差而进行平滑和插值的结果。请注意发生在实验的"死时间"内没有变化,它表明在这段时间内没有形成任何中间体。图片经许可取自 Kragelund B. B., Robinson C. V., Knudsen J., et al. 1995. Biochemistry. 34: 7217-7224. © 1995, ACS

■ 然而,在这些实验中是否可以进行研究,以阐明自动折叠过程的本质?因为里面没有可以隔离并研究的中间体?!

回答是:在这里可以研究**过渡态**,在蛋白质折叠路径中"最狭窄的位置"。

我想,应该从最开始就指出,在折叠过程的动力学中着关键作用并决定其速度的过渡态,顾名思义,是在折叠路径中完全不稳定的,**最不稳定**的状态。因此,这**不是**任何一种上面讨论到的"折叠中间体"(尤其是,**不是**熔球!)。"折叠中间体"对应着自由能极小值;因此它们至少在某个时间段中是稳定的,即它们可以在折叠过程中积累起来并可以被直接观察到。相反,过渡态对应着自由能极大值,因此它不能积累也不能直接观察到,只能根据它对折叠速度的影响而观察到。

■ 让我们暂且推迟详细介绍对在蛋白质折叠中观察到的过渡态的研究,而是跑到前面,仅仅先总结它的主要结果。

在过渡态研究中勾勒出了蛋白质折叠晶核,"折叠核"(图19-5,参见彩色立体图对D-29和D-30)。在这个晶核中链上相对少量的残基已经具有其天然的(即与天然蛋白质中相同的)空间排布和构象,与此同时其他残基仍然停留在变性相。

所有这些我们将在下一课中更详细地讨论。

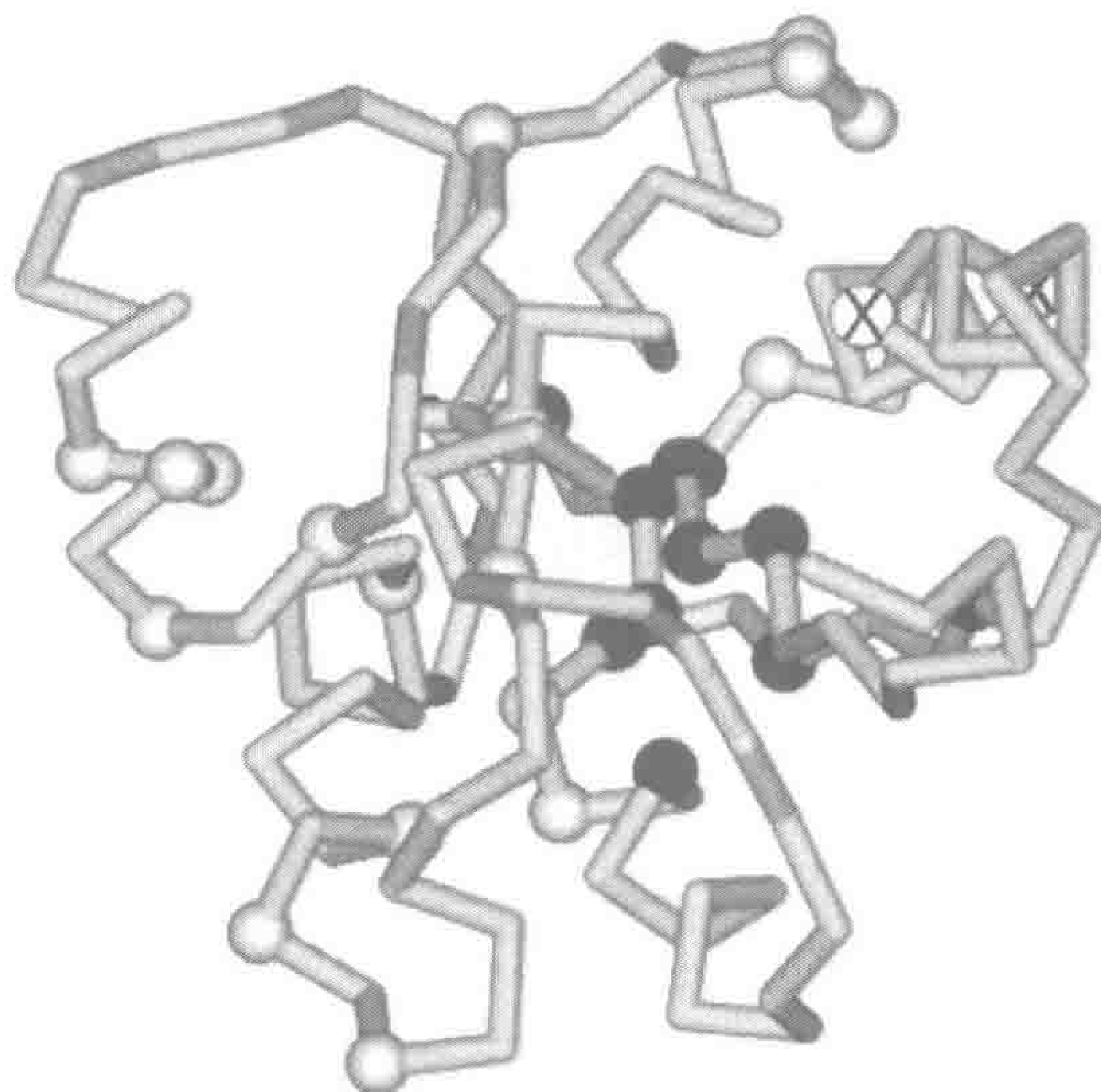


图 19-5　通过实验勾勒出的 CheY 蛋白质的折叠核，据 López-Hernándes，Serrano. 1996. Folding and Design. 1：43-55。以肽链的天然折叠为背景，进入到折叠核的残基以黑圆球表示。灰色圆球表示未参与到折叠核的残基。无球的位置表示尚未做过突变的肽链区域。“×”号标出的是两个难以实验解释的残基

■　到目前为止，我们只讨论了水溶性球状蛋白。

事实上，对蛋白质折叠的研究绝大多数都是在这类蛋白质上进行的。但是在每一种生物体的基因组都有三分之一的蛋白质是膜蛋白，它们只能在有脂类存在的条件下才能达到的自身的功能状态。目前，有一些有利的数据，说明螺旋形膜蛋白的折叠是通过中间态进行的，在很多方面与在水溶性球状蛋白中观察到的熔球相似，即具有类天然的二级结构，但缺少空间结构中的相互作用有序性。从紧密熔球上有疏水残基伸出，这使得它们在细胞膜内折叠时非常吸引人。有一些试验使人们认为，在膜蛋白中的不同螺旋或多或少自动地折叠，这与对熔球预期的柔性相一致。在膜蛋白和水溶性蛋白的折叠之间的相似性可以推知，可能存在有可能作为其折叠晶核的结构性“小核”。其他膜蛋白，如具有 β 桶型结构的膜孔蛋白，显然比螺旋形膜蛋白更协调得多地折叠，即在它们的折叠过程中看不到清晰的阶段和稳定的中间体。

■　在蛋白质折叠研究领域的进展，除了满足我们天生的好奇心之外，毫无疑问，还必将有助于我们有效地表达水溶性球状蛋白和膜蛋白的能力，并增加我们对细胞内过程、分泌过程和膜过程的调控的理解。

■　在本讲的结尾处，讨论一下蛋白质的自动折叠和病毒的自动组装之间的某些

相似性并非多余。在许多简单病毒的外壳蛋白中，看上去具有对形成完整的病毒微粒必不可少的全部性质。然而对于更复杂的病毒，已知它们的组装过程中包含有“脚手架”，即蛋白质或其他对形成微粒必需的、但不包括在成熟的病毒粒子中的成分。无论是从结构的观点，或者从组装机制的观点来看，这些“分子伴侣”都很值得注意。有趣的是，由于简单病毒的组装过程中没有中间体的积累（恰好就如简单蛋白质折叠过程中没有部分折叠结构的积累一样），因此与简单病毒（和蛋白质）的基本组装路径相比，我们对复杂病毒（蛋白质也是如此）的更复杂的组装路径积累了更多得多的信息，因为部分地已形成结构的粒子可以在适当条件下可以被观察到。

第 20 讲

小蛋白质的单步骤折叠。过渡态理论。在实验上寻找和研究蛋白质折叠中的不稳定过渡态。天然蛋白质结构的折叠核。蛋白质工程上用体外方法对它们的实验观察。蛋白质折叠的成核机制。

如上一讲所述,在了解蛋白质折叠中,对蛋白质自动折叠的最简单情形(折叠过程中不发生任何亚稳中间状态积累)的研究发挥了最重要的作用。

我提醒大家,某些小蛋白质的折叠(在体外模拟的"生理"条件下)速度很快,在实验中**没有**观察到中间体的"积累"。因此这种转变被认为是仅发生于一个步骤:这是根据蛋白质中的二级结构、侧链基团的拼装、色氨酸荧光(依赖于色氨酸埋入到球内的程度)和氢交换率等性质的恢复速度都完全相同而确认的。

值得强调的是,已知单域蛋白质的变性和复性热力学对应着"全或无"型转变,即缺乏任何明显数量的"半折叠"形式,由此可以先验地(*a priori*)预期,单域蛋白质在变性/复性点附近进行的折叠是单级动力学过程。不寻常的是,对于一些小蛋白质,在远离平衡的"生理"条件下的折叠也是如此:在这种条件下它们的折叠速度特别快,而大部分其他蛋白质折叠过程中有熔球型中间体的积累。

正是在这种在广泛的外部条件范围中发生(且没有复杂的对其中间体积累的分析)的单步骤折叠中,最方便研究**过渡态**在过程动力学中起着关键作用的状态。

■ 回忆一下,过渡态对应着从一个稳定状态(自由能**最小**)到另一个稳定状态的路径上的自由能**极大值**。

还提醒一下,在两个直接被观察的(稳定的)状态(图 20-1 中的 A 和 B)之间的最简单转变的动力学,可以成功地用**过渡态理论**所描述,这一理论在化学中通常被称为"**活性复合物理论**"。

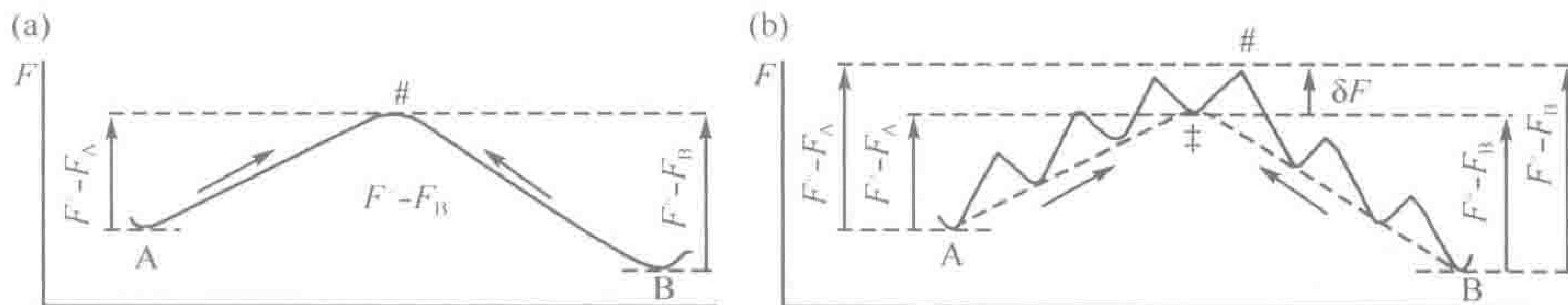


图 20-1 从稳定态 A 到稳定态 B(从左向右箭头)和从 B 到 A(从右向左箭头)的转变过程中的自由能垒的克服。F_A、F_B 和 $F^{\#}$ 表示状态 A、B 和"过渡态"#(在途径中具有最高自由能的能垒态)。(a)单元反应。(b)由几个单元反应组成的、穿越具有较高自由能(高于 A 和 B)中间状态的过程。δF 为在单个步骤中需要克服的自由能升高量。‡ 为稳定状态中具有自由能最大值 $F^{\ddagger}$ 的中间态

过渡态理论的本质归结为，过程的速度受到**过渡态#**，即具有**最小**（根据热力学）**概率**（因此不积累也不能直接观察到）、**但**从 A 到 B 的路径中**所必须经过的中间状态**，或者，换句话说，即 A→B（和 B→A）的转变动力学取决于连接 A 和 B 的路径上的自由能最大值的**高度**。

提醒一下，因此从 A 到 B 和从 B 到 A 的转变速度可按下式描述：

$$k_{A\to B}=k_0\ \exp[-(F^{\#}-F_A)/RT]$$
$$k_{B\to A}=k_0\ \exp[-(F^{\#}-F_B)/RT] \qquad (20.1)$$

式中，F_A、F_B 和 $F^{\#}$ 为状态 A、B 和"过渡态"#（图 20-1）的自由能，而 k_0 为该过程的"单元步骤"的速率。

在讨论单元反应[图 20-1(a)]时，k_0 通常取值为 $k_{0,el}=RT/h$（其中 h 为普朗克常量），$T\sim300$K 时约等于 $10^{13}\ s^{-1}$。（这是分子在热涨落的作用下试图克服活化能垒 # 的频率。）

然而，如果假设它们的自由能比 F_A 和 F_B 高得多，则我们可以更方便地[图 20-1(b)]只考虑稳定状态（在反应路径中的自由能局部最小值）。我提醒大家，这种多步过程的速率（如果步数不是太多）也仅取决于其路径中的最高能垒。

在考虑含有多个步骤的过程时，方便的是：

（1）单个步骤的速率设为等于从一个最小值到相邻的另一个最小值的转变速率

$$k_{0,step}=k_{0,el}\ \exp(-\delta F/RT)$$

（式中，δF 为在单个步骤过程中克服的自由能升高）；

（2）在自由能垒中 $F^{\ddagger}$ 取为在该过程中的稳定中间态的最大自由能。则有

$$k_{A\to B}=k_{0,el}\ \exp[-(F^{\#}-F_A)/RT]=k_{0,el}\ \exp[-(F^{\ddagger}+\delta F-F_A)/RT]$$
$$=k_{0,step}\ \exp[-(F^{\ddagger}-F_A)/RT] \qquad (20.2)$$

式(20.2)与式(20.1)相似，作替换 $k_0\to k_{0,step}$，$F^{\#}\to F^{\ddagger}$ 即可。

因此 A→B 转变的特征时间为

$$t_{A\to B}=\frac{1}{k_{A\to B}}=\frac{1}{k_{0,step}}\exp[+(F-F_A)/RT] \qquad (20.2a)$$

这些说明很自然地可以推广到 $t_{B\to A}$ 和 $k_{B\to A}$。

根据这些校正，式(20.1)可以应用到符合如下条件的过程：①在比单元步骤时间长得多的时间内进行；②由并不太多的步数组成；③不具有稳定的（发生积累的）中间体。而我们在蛋白质折叠中遇到的正是这样的过程。

■ 重要的是，A→B 的反应速度对温度的依赖性可以用于判断过渡态的能量。根据 Arrhenius 公式①，计算出反应速度的对数值对温度倒数（$1/T$）的导数：

---

① Arrhenius（阿累尼乌斯）公式用于描述化学反应常数 k 与温度 T 及活化能 E_A 之间的关系：$k=Ae^{-E_A/RT}$，其中 R 为理想气体常数。

$$\frac{d(\ln k_{A\to B})}{d\left(\frac{1}{T}\right)}=-\frac{1}{T^2}\frac{d\left[\ln k_{0,\mathrm{el}}-\frac{(F^{\#}-F_A)}{RT}\right]}{dT}\approx-\frac{E^{\#}-E_A}{R} \tag{20.3}$$

此处我使用了大家熟悉的公式 $\frac{d(R/T)}{dT}=-\frac{E}{T^2}$，忽略了单元步骤速率 $k_{0,\mathrm{el}}$ 对温度的极微弱的依赖性。后者甚至在简单的化学反应中也可以接受，所以在蛋白质中也同样可以接受：在蛋白质中由于能量 $E^{\#}$ 和 E_A 是由大量粒子的相互作用所决定的，所以都很大。

■ 图 20-2 显示了溶菌酶(lysozyme)的折叠和解折叠速率在热变性和复性中的变化。

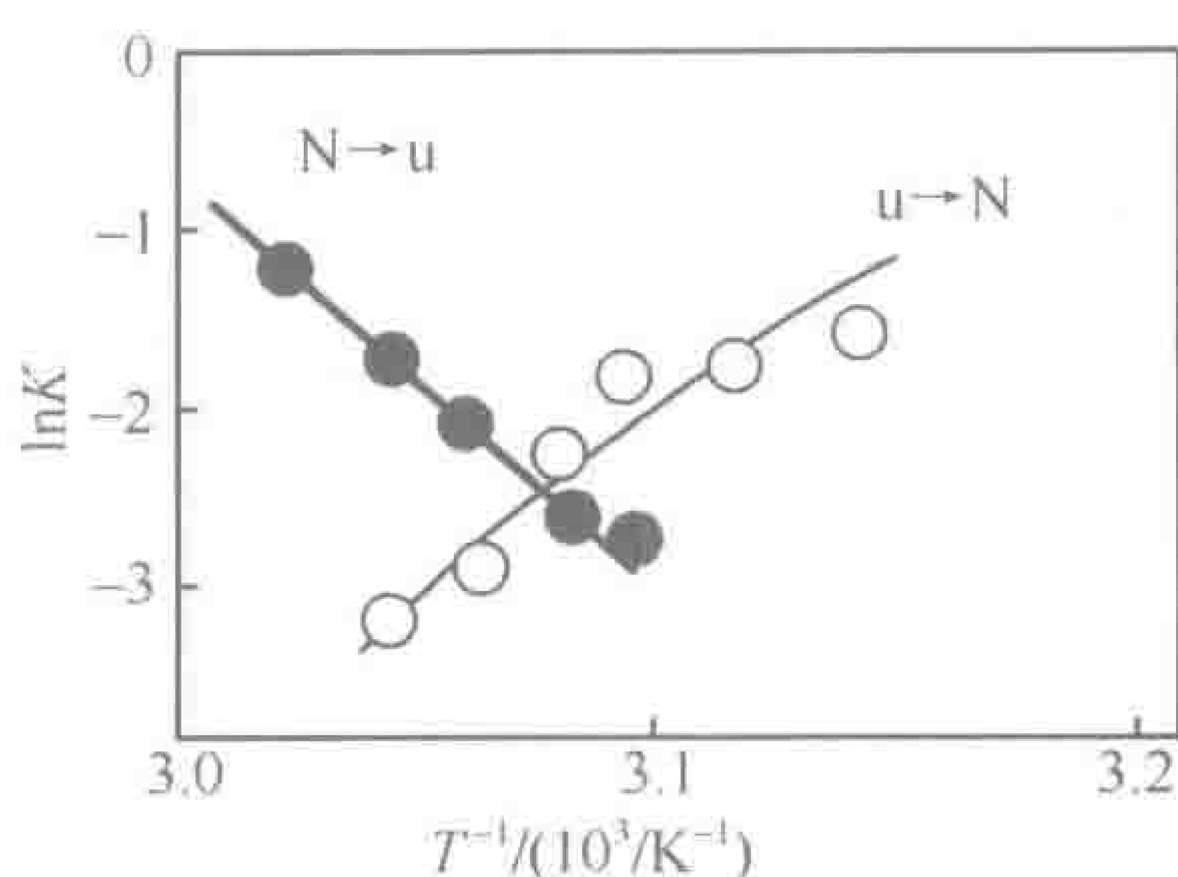


图 20-2 溶菌酶的热变性和复性速率与温度倒数 T^{-1} 的 V 形关系曲线图；图片取自 Segava S. , Sugihara S. 1984. Biochemistry. 23：2473-2488。速度常数 k 以 s^{-1} 为单位。复性速率为 $k_{u\to N}$（○点为实验值，细拟合曲线）；变性速率为 $k_{u\to N}$（●点为实验值，粗拟合曲线）。熔化中点对应于 $k_{u\to N}=k_{N\to u}$ 时的温度，即曲线相交处（约为 1000/3. 08K = 325K）。在温度更低（即 T^{-1} 更大：从相交点向右）时折叠占优势，在温度更高时（即 T^{-1} 更小：从相交点向左）解折叠占优势

图中显示，在转变正中央处［在该处折叠速率 $k_{u\to N}$ 等于解折叠速率 $k_{N\to u}$，因此曲线 $\ln(k_{u\to N})$ 与 $\ln(k_{N\to u})$ 相交］这两个过程的速率接近于 $e^{-2.5}\approx0.1\ s^{-1}$（即此处变性和复性的进行时间约为 10s）。

除此之外，该图还显示，通过深入到“变性区域”可使变性加速，而深入到“复性区域”可使复性加速。而且这里含有一个非常有趣的结果。

情况是这样：热变性速率 $k_{N\to u}$ 随 T^{-1} 而**降低**（即随温度 T 而**升高**，物理化学反应是通常如此），而折叠速率 $k_{u\to N}$ **相反**，是随 T^{-1} 而**升高**（即随温度而**降低**，物理化学反应通常并**不**如此）。根据式(20.3)，这意味 $E^{\#}-E_N>0$ 而 $E^{\#}-E_u<0$。换句话说，$E_u>E^{\#}>E_N$，即能垒能量 $E^{\#}$ 高于天然态能量 E_N，但低于解折叠态能量 E_u（后者对化学反应是异常的，化学反应中能垒能量**既高于**反应初态，**也高于**反应末态）。对此图的进一步分析［采取 $T\ln(k/k_{0,\mathrm{el}})$ 与 T 的关系曲线形式］表明，这种关系对于蛋白质折叠中的解折叠态、过渡态和天然态的熵也是适用的：$S_u>S^{\#}>S_N$。这意味着，如

果从天然态方向看过去的话，蛋白质的天然态和解折叠态之间的能垒像是“寻常的”能量能垒，而如果从解折叠态方向看过去，则像是异常的。

不过，从上一课中已经做过的自由能垒分析和蛋白质天然态与变性态之间“全或无”型相变原因分析出发（图 20-3，该图重复了你们已经熟悉的能量、熵和自由能随熔球膨胀而变化的示意图），这是可以预期的。

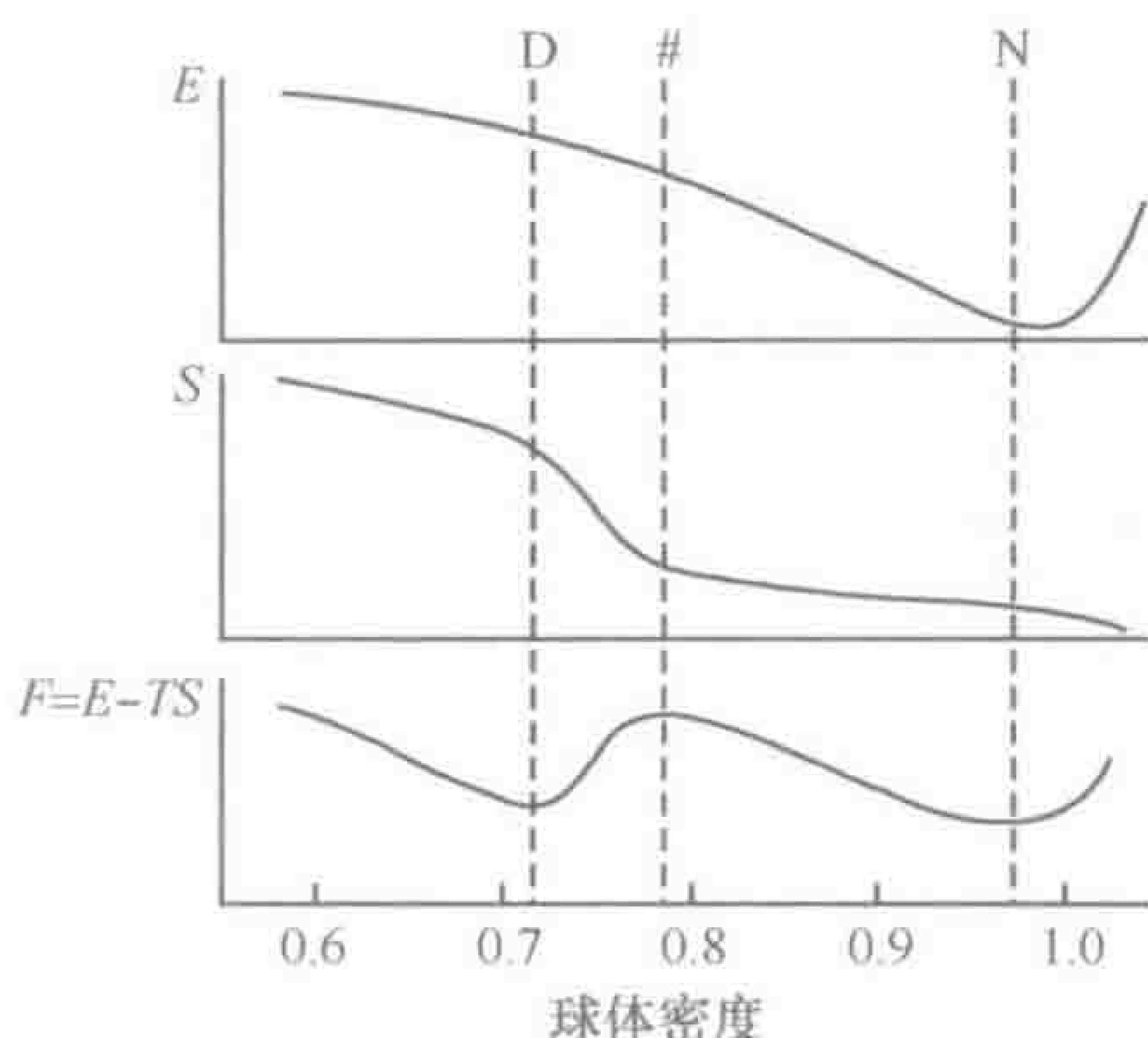


图 20-3　能量 E、熵 S 和自由能 $F=E-TS$ 随球体密度的变化。D：变性态（本例中为熔球，因为其密度较大）；N：天然球体；#："能垒"，球体均匀膨胀过程（从 N 到 D）中的自由能最大值

■　回头再来看方程组（20.1）。它们表明，正反应和逆向反应的速率之比，$k_{A\to B}/k_{B\to A}$，就是末态 B 与初态 A 之间的平衡常数：

$$K_{B:A}=k_{A\to B}/k_{B\to A}=\exp[-(F_B-F_A)/RT] \tag{20.4}$$

众所周知，这个平衡常数的值表明了处于状态 B 的最终分子数目 n_B^∞ 与处于状态 A 的最终分子数目 n_A^∞ 之比。即过程在给定条件下（温度等）最终计数（在观测时间趋于无穷大时）时，即 A→B 流等于 B→A 流，以致 $n_A^\infty k_{A\to B}=n_B^\infty k_{B\to A}$时所达到的值：

$$K_{B:A}=n_B^\infty/n_A^\infty=k_{A\to B}/k_{B\to A} \tag{20.4a}$$

系统会以怎样的速度进行到这个平衡状态？为此让我们来求解相应的微分方程：

$$\frac{dn_A}{dt}=-k_{A\to B}n_A+k_{B\to A}n_B \tag{20.5}$$

式中，t 为时间，而 n_A 和 n_B 为分别处于状态 A 和 B 的分子数（第二个方程 $\frac{dn_B}{dt}$ 不需要求解，因为 $n_A+n_B\equiv n_0$，其中 n_0 为全部分子，即有 $\frac{dn_B}{dt}=-\frac{dn_A}{dt}$）。

把 $n_B=n_0-n_A$ 代入到式（20.5）中，并利用 $\frac{n_0k_{B\to A}}{k_{A\to B}+k_{B\to A}}=n_A^\infty$，根据式（20.4a），得到

$$\frac{dn_A}{dt}=-(k_{A\to B}+k_{B\to A})(n_A-n_A^\infty)$$

最后结果是（请允许我不列后续计算而写出，但建议你们自己做一下）这样：

$$n_A(t) = [n_A(0) - n_A^{\infty}]\exp[-(k_{A\to B} + k_{B\to A})t] + n_A^{\infty} \quad (20.6)$$

式中，$n_A(t)$为从过程开始后t时刻处于状态A的分子数，而$n_A^{\infty} = n_0 \frac{k_{B\to A}}{k_{A\to B} + k_{B\to A}}$，为处于状态A的最终的平衡分子数目。

这意味着，接近平衡时的表观速率为

$$k_{app} = k_{A\to B} + k_{B\to A} \quad (20.7)$$

它等于正反应和负反应的速率之和。

请注意，k_{app}这个速率只取决于反应进行的条件，而不依赖于蛋白质分子在天然态和变性态的初始分布（如初始分率$n_A(0)$）。

在总速率k_{app}中占优势的是速率更快的反应。如果条件（温度等）更适合于天然态的稳定性，则在$k_{app} = k_{u\to N} + k_{N\to u}$占优势的是折叠速率$k_{u\to N}$。如果变性态更稳定，则在$k_{u\to N} + k_{N\to u}$中占优势的是解折叠速率$k_{N\to u}$。

无论是天然态更稳定，或是变性态更稳定，接近平衡时的速率k_{app}总是可以被测量的（与单独的$k_{u\to N}$和$k_{N\to u}$不同），这是它的一个重要优点。

■ 通过测量k_{app}可以绘制出所谓“V形图”（图20-4）。这样称呼是由于其V字形的形状让人想到士兵短袖上的袖章①。这种图中一眼可见，折叠速率$k_{u\to N}$和解折叠速率$k_{N\to u}$与变性剂浓度的依赖关系刚好相反。我们已经在研究蛋白质的折叠和解折叠速率对温度的依赖关系时碰到过类似的情形。与那里相同的是，此处观察到的斜率的相反性意味着过渡态性质的“中间性”，它们处于天然蛋白质与解折叠蛋白质的中间位置。

确实，变性剂总是使蛋白质解折叠，这表明，它对解折叠蛋白质的吸引比对天然蛋白质的吸引更强烈。而变性剂能减慢初始解折叠蛋白质的折叠（图20-4左边部分的速率降低），说明它对初始解折叠蛋白质的吸引比对过渡态的吸引更强烈；而变性剂能加速初始天然蛋白质的解折叠（图20-4右边部分速率的上升），表明它对过渡态的吸引要强于初始天然蛋白质。这意味着，变性剂与过渡态的接触强于与天然蛋白质的接触，但弱于与已变性蛋白质的接触。换句话说，图20-4中的V形图表明，根据其与溶剂的接触程度，即自身的紧密性，过渡态处于蛋白质的解折叠态和天然态的半途中的某处。该图还说明，因为$k_{N\to u}$的斜率比$k_{u\to N}$要小，则过渡态的紧密性比较接近（在本例中）蛋白质天然态，而非变性态。

图20-4（以及图20-2）所示的是溶菌酶的情况，这种蛋白质由盐酸胍引起的变性直接导致形成线团，而不是熔球。在适当的变性剂浓度（没有被水强烈稀释）下，溶菌酶折叠得相当慢，需要好几分钟。然而在几乎纯净的水中自动折叠（用水强烈稀释含有线团状蛋白质链的盐酸胍浓溶液）时，溶菌酶的折叠速率最大，且只微弱地依赖于剩余的变性剂浓度。在这种情形下溶菌酶的折叠中显示出紧密的熔球型亚稳中间体，这种中间体在较高浓度的变性剂中未被观察到。

① 原文直译为“袖章图”，为方便理解而译为“V形图”。

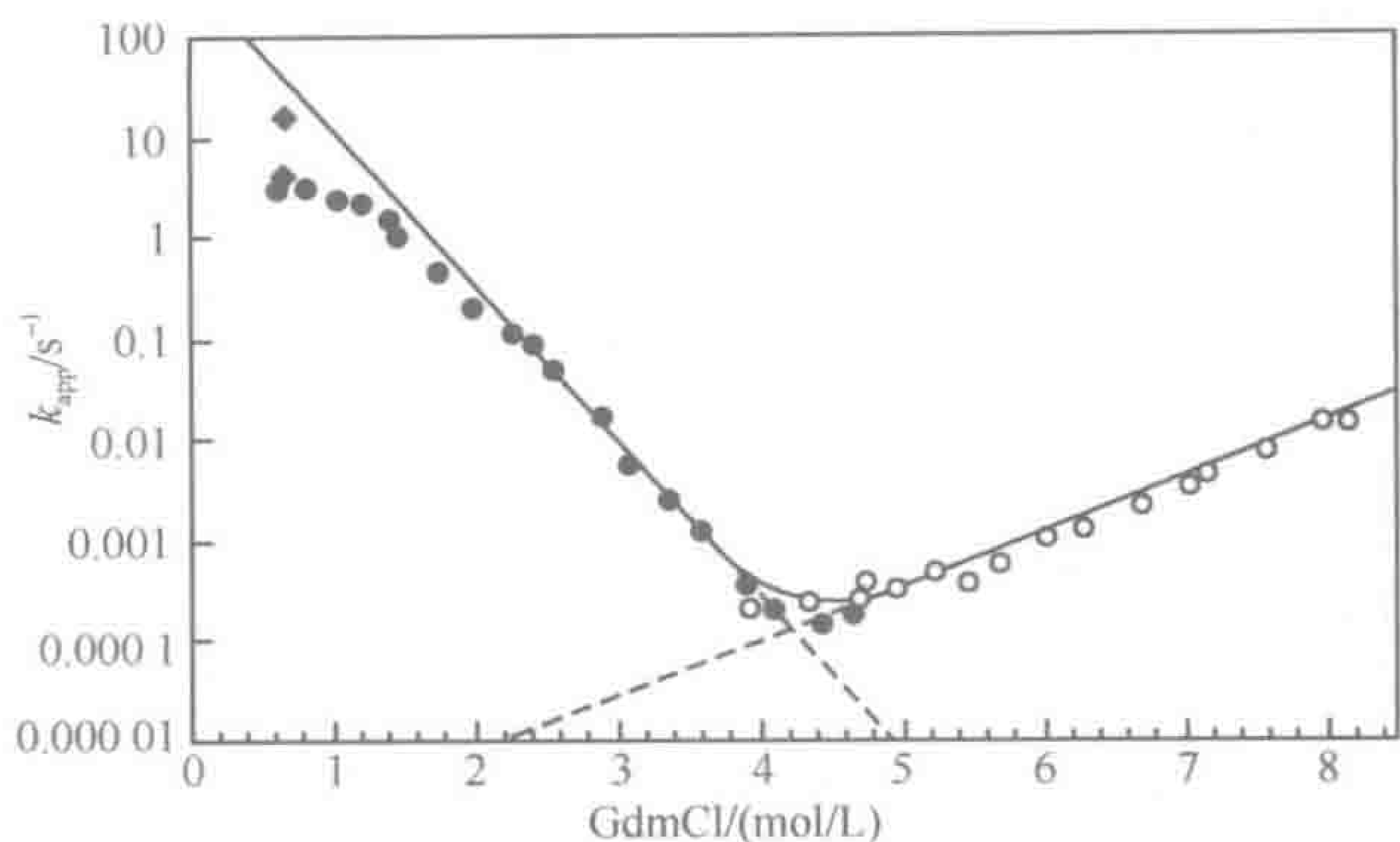


图 20-4　蛋白质(鸡卵清溶酶菌)天然形态和解折叠形态之间接近平衡时,表观特征速率 $k_{app}=k_{u\to N}+k_{N\to u}$与盐酸胍浓度关系曲线的"V 形图"。实验是在存在有天然二硫键的情况下进行的。实心圆点是在稀释 GdmCl 浓溶液(开始时变性蛋白质置于其中)(即蛋白质复性)过程中得到的。因此 $k_{app}\approx k_{u\to N}$(因为这里折叠速率 $k_{app}\approx k_{u\to N}$远大于解折叠速率 $k_{N\to u}$)。空心圆点是在开始时为天然蛋白质的溶液中添加 GdmCl(即蛋白质变性)的过程中得到的。因此 $k_{app}\approx k_{N\to u}$(因为在变性过程中 $k_{u\to N}<<k_{N\to u}$)。在 V 形曲线的最下方部分 $k_{u\to N}\approx k_{N\to u}\approx k_{app}/2$。请注意实心圆点和空心圆点在这一区域(即变性点附近)相交:这表明,蛋白质在此处的折叠和解折叠的速率确实相同。虚线表示 $k_{u\to N}$和 $k_{N\to u}$在 V 形曲线的拐点处的外推,而在其之上,实线表示在低浓度 GdmCl 区域的拟合。在图中左上方偏离拟合直线的点(即**远离**拐点:蛋白质变性点)说明可能是出现过渡态重建,或出现任何额外的可能处于主要折叠路径之中或之外的亚稳中间体(有可能是熔球)。请注意,所有这些重建和/或中间体都**不能**增大折叠速度以超过与过渡态保持不变时所能期望的速率;偏离点尽管对应于观察到的最高折叠速率,但仍然位于外推直线**下方**。图片取自 Kiefhaber T. 1995. Proc Natl Acad Sci USA. 92:9029-9033,略有修改(增加了外推虚线)

■　因此,远离平衡点时溶菌酶显示出两级折叠,接近平衡时显示为单步骤折叠,即如果仅在接近平衡点处讨论,则在"两步骤"折叠和"单步骤"折叠的蛋白质之间没有物理差别。

然而,研究"单步骤"折叠的蛋白质更简单:它们在水中的过渡态与在所有变性剂的不同浓度下的过渡态都是一样的。而且还有技术上的优势:V 形图的长分支更容易进行外推。因此正是对"单步骤"蛋白质的折叠研究在过渡态的研究中起着主导作用。

■　一般来说,紧密亚稳折叠中间体的出现,总是使折叠速率与变性剂含量的关系曲线"扁平化"(图 20-4 左边区域),甚至改变其斜率的正负号,因为限制折叠速率的步骤,现在是从更紧凑的状态开始。与原来开始时的解折叠态相比的话,其性质(如与变性剂的结合性)更接近于过渡态的性质。已经知道,从熔球而非线团开始的蛋白质复性(如碳酸酐酶的复性)也显示出复性对变性剂

浓度的依赖性相当弱。这意味着，在这种折叠中的过渡态在紧密性方面与初始熔球差别很小。

换句话说，过渡态（根据其某些性质在天然蛋白质和解折叠蛋白质之间的"中间性"）与熔球相似。在能量和紧密性这些"平均"特征方面相似，**但**这并不意味着过渡态（强调一下，它是不稳定的，与熔球不同）真的像熔球。

实验（我将在几分钟后谈到）显示，过渡态远比熔球不均匀得多。它（如果讨论的是"线团→天然蛋白质"转变）可能是天然蛋白质小块，其余部分仍然处于变性的线团形状态。"熔球→天然蛋白质"（以及"线团→熔球"）转变中的过渡态的外观在实验上仍然没有建立起来，但很可能的是，它们也是由较多结构形成相的小块组成，与此同时肽链的其余部分仍然停留在较少结构形成相。

■　借助于蛋白质工程，过渡态的本质已经被成功地阐明。通过进行大量突变，并分析它们在 V 形图上相应的变化（图 20-4，图 20-5），这样花费了巨量的劳动之后终于了解到，哪些残基参与到过渡态的"类天然"部分（其突变严重影响折叠速率），而哪些残基没有参与。这一方法是英国的 A. Fersht①发展起来的。

这种方法适用于任意"全或无"型转变。然而，严格来说，至今它只应用于那些直接变性到形成线团而非熔球的蛋白质。

■　为了评价残基在过渡态的类天然部分（或者说，在蛋白质的"晶核"或"折叠核"）中的参与程度，可以根据 V 形曲线的移动，估算给定残基的突变对①蛋白质折叠速率和②蛋白质天然形状的稳定性的影响。为研究折叠核而对残基进行突变时，通常突变为更小的残基[常见的是丙氨酸（Ala）或甘氨酸（Gly）]，以便突变后的蛋白质结构不会变形太大。测量方法显示在图 20-5 中。

蛋白质天然结构 N 相对于其解折叠态 u 的稳定性（在给定的环境条件下，如给定变性剂浓度）决定于该蛋白质分子的这两个状态的自由能（在相同的条件下测量）之差 F_N-F_u 的值。而 F_N-F_u 决定于（式 20.4）蛋白质在给定的环境条件下的折叠速率与解折叠速率之比。由于蛋白质的稳定性通常是相对纯水而言的，因此在高浓度变性剂下测量得的解折叠速率通常需要外推到其低浓度条件。

蛋白质折叠速度（u→N）决定于（式 20.1）蛋白质的"晶核"# 与初始的解折叠态 U 之间的自由能之差（即 $F^\#-F_u$）。

突变也会影响折叠速率和蛋白质的稳定性。这使得它可以用于在实验上测定蛋白质天然结构的晶核形状。

在对折叠的**成核机制**的设想中，要解释突变对 $F_N - F_u$ 与 $F^\# - F_u$ 之比的影响，归结为，残基（如果它们参与到天然蛋白质晶核中的话）所处位置是否与在天然蛋白质中相同？

---

① Alan Roy Fersht（1943—），英国剑桥大学教授，在蛋白质折叠研究方面作出了先驱性的工作。

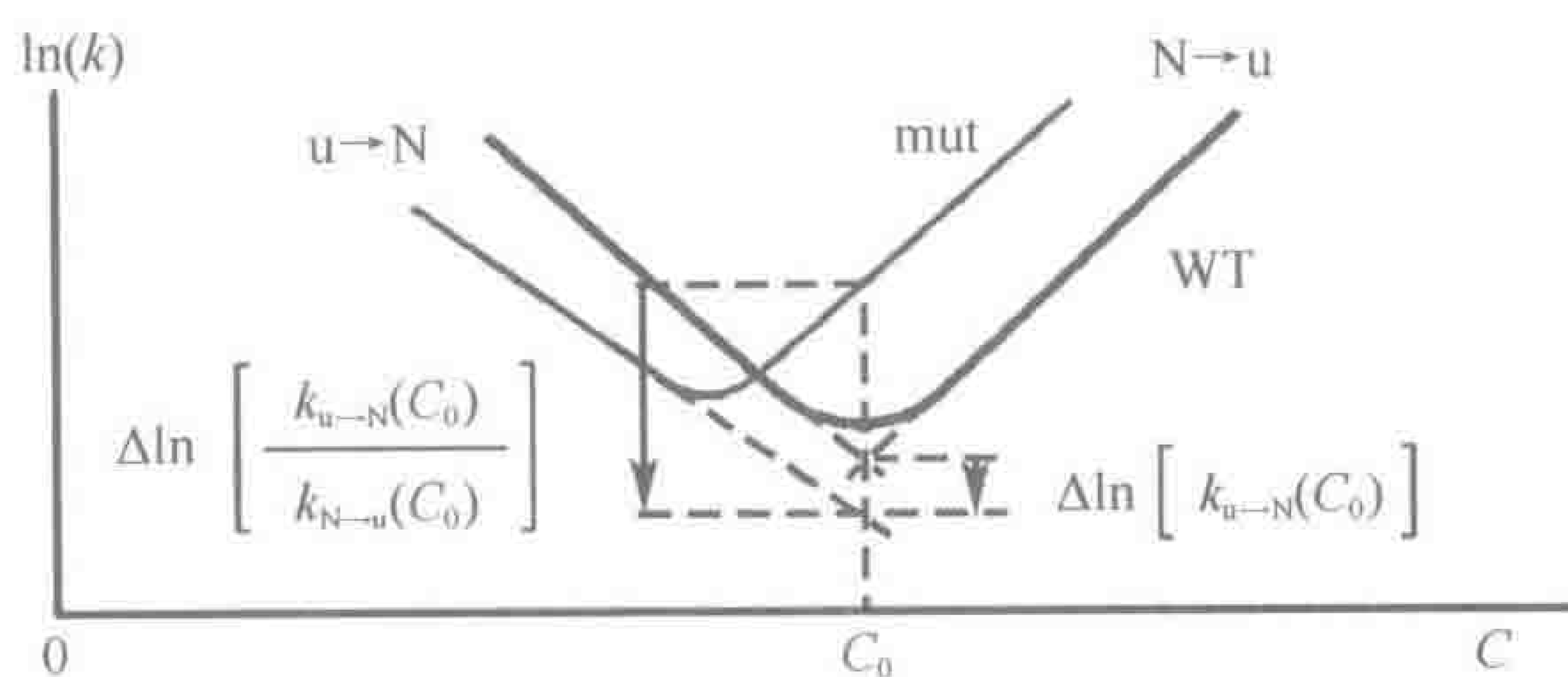


图 20-5　说明突变时"V 形图"移动的示意图。k 为观察到的平衡建立速度；C 为变性剂浓度。粗实线为野生型蛋白质(wild type,WT)，细实线为链上一个残基突变的同一蛋白质。虚线表示 $k_{u\to N}$ 和 $k_{N\to u}$ 值在 V 形曲线拐弯处的外推。C_0 为野生型蛋白质变性过程的中点所对应的变性剂浓度(根据"转变中点"的定义，此处天然蛋白质和变性蛋白质的自由能之差等于零)。图中还显示了在此实验中测得的值。其中一个能确定变性对从解折叠态到天然态的转变路径中的自由能垒$[\Delta(F^{\#}-F_u)=-RT\Delta\ln k_{u\to N}]$的影响。另一个确定对蛋白质稳定性，即天然态与变性态的自由能之差：$\Delta(F_N-F_u)=-RT\Delta\ln\frac{k_{u\to N}}{k_{N\to u}}$ 的影响。两者之比 $\frac{\Delta(F^{\#}-F_u)}{\Delta(F_N-F_u)}$ 被称为被突变残基的 Φ_f 值。在此图中 $\Phi_f\approx1/4$，即所讨论的残基有 1/4 的程度参与到折叠核。图中所显示的值与需要最小外推(因此其误差也最小)的变性剂浓度 C_0 有关。然而，通常外推更多(至 $C=0$)，就能确定在纯水中的天然蛋白质稳定性变化 $\Delta(F_N-F_u)$ 和能垒变化 $\Delta(F^{\#}-F_u)$ 以及 Φ_f 值

在蛋白质折叠中存在有成核作用是非常合理的，因为这种折叠是以"全或无"型转变(而我们已经知道，这是一级相变在宏观系统中的微观相似物)进行的。成核作用对一级相变(如结晶)是很典型的。

我会很快就讲到这一设想的实验验证。首先我想讲到，F_N-F_u 和 $F^{\#}-F_u$ 是如何利用来描画出天然结构的晶核("折叠核")的。

■　如果残基突变既影响天然态的稳定性(即影响 F_N-F_u 的值)，也同样地影响 $F^{\#}-F_u$ 的值，则说明所研究的残基既参与到天然球体中，也参与到"晶核"中(在那里形成接触，具有相同的构象等)。

相反，如果残基突变只影响蛋白质的稳定性(即影响 F_N-F_u 的值)，而**不**影响折叠速率(即它**不**影响 $F^{\#}-F_u$ 的值)，则意味着该残基**未**参与到球状的折叠核，即它只是在晶核形成之后才进入天然蛋白质球体。

最后，如果残基突变强烈地影响蛋白质的稳定性，而较弱地(以相同的正负号)影响晶核的稳定性，则意味着，该残基要么是进入到几个可选的晶核之一，要么在晶核中只形成在天然蛋白质中所具有的部分接触，即处于晶核表面。

蛋白质折叠晶核就这样被描画出来(图 20-6，参见彩色立体图对 D-29 和 D-30)：对每个突变残基计算如下数值：

$$\Phi_f=\frac{\Delta(F^{\#}-F_u)}{\Delta(F_N-F_u)} \tag{20.8}$$

而且,如果给定残基的 Φ_f 值接近于 1,则说它进入到结构晶核中,如果 Φ_f 值接近于 0,则没有。

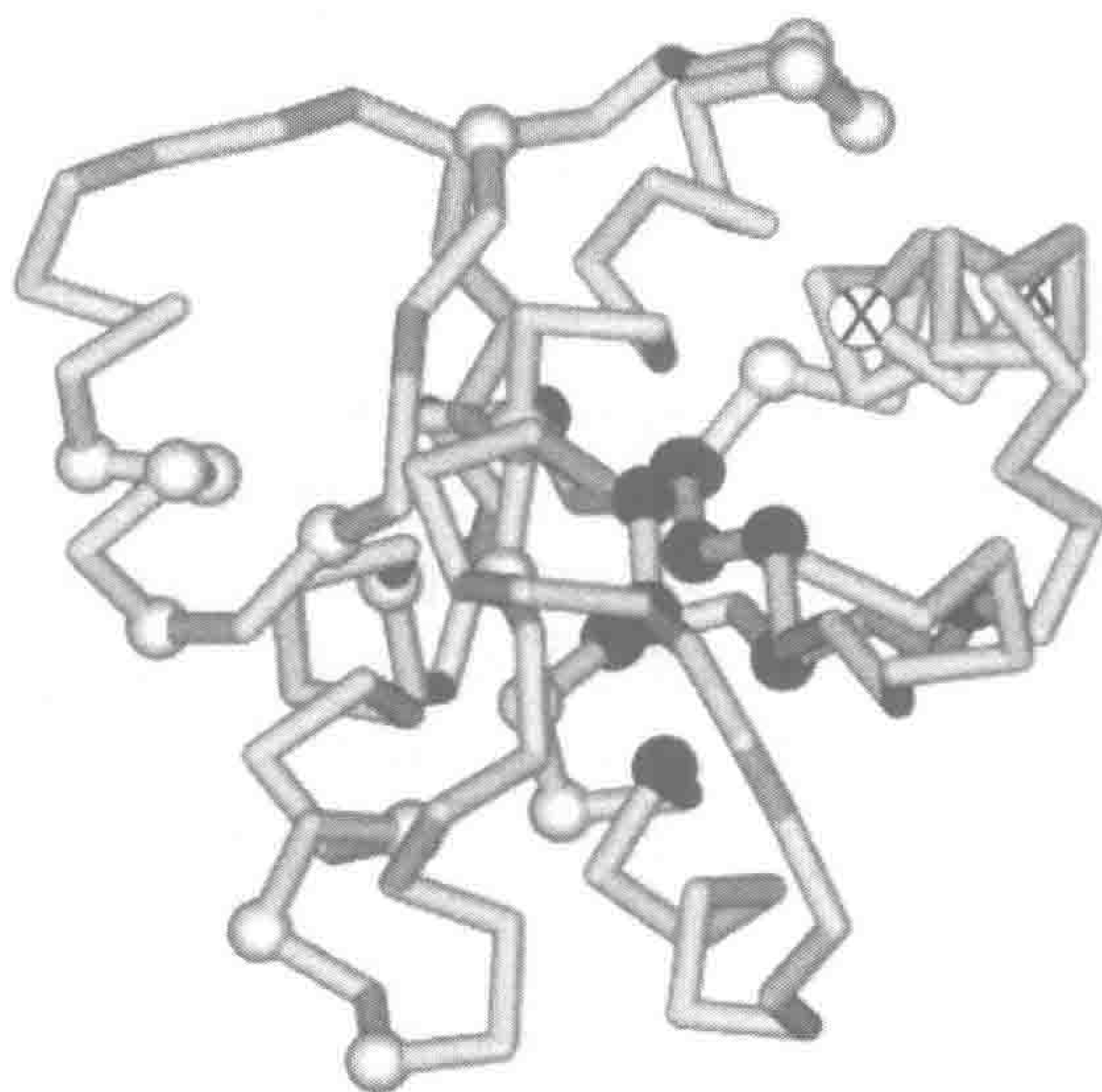


图 20-6 蛋白质 CheY 的过渡态结构,据 López-Hernándes, Serrano. 1996. Folding and Design. 1: 43-55。在该链的天然折叠背景上,参与到过渡态中已形成结构部分的残基以黑圆球表示(具有 Φ_f>3,即那些位置已形成了 30% 以上的天然接触),更明亮的圆球表示未参与到过渡态已形成结构部分、Φ_f<3 的残基。其余无球的部分表示尚未做过突变的肽链区域。"×"号标出的是两个难以实验解释的残基,它们的 $\Delta(F^{\#}-F_u)$ 和 $\Delta(F_N-F_u)$ 值太小($\ll kT$)(后者更重要,因为该值位于式(20.8)的分母中),以致测定时的误差超过了其值本身

值得注意的是,蛋白质中只有少量残基不能用这一观点解释。以这一观点**不能**解释那些未参与到折叠"成核机制"中的残基,即那些突变后只影响折叠速率而不影响天然结构稳定性的残基,或那些只使晶核稳定而使天然蛋白质不稳定的残基,等等。在这些残基上得到了 Φ_f>1 和 Φ_f<0 的值,这极其少见,而且通常只是在 $\Delta(F^{\#}-F_u)$ 值测量不准确且 $\Delta(F_N-F_u)$ 值特别大的情况下。

这让人们确信,基础图景(参与晶核的残基处于与在天然蛋白质中相同的位置)基本上是正确的。我要着重指出,是基本上,而非完全地。有一些数据显示,处于"过渡态"的蛋白质链比线团环绕的"类天然晶体"所能预期的更紧密。可以谨慎地设想,至少在某些蛋白质中,过渡态的"类天然晶体"是被某种熔球所围绕。

■ 图 20.6 表明,在蛋白质 CheY 中按上述方法找到的对该蛋白质折叠最为重要的残基,聚集为紧凑的"小域",即紧凑的折叠核,而且它不是位于蛋白质的几何中心,而是靠近其表面。在其他用于作为折叠核位置研究对象的蛋白质中也常常观察到了类似的图景。这表明,蛋白质的折叠核(folding nucleus)与其疏水核(hydropobic core)是不同的东西。在其他一些蛋白质中观察到了非常"发散的"折叠

核,涵盖蛋白质的大部分、但弱表达(特征为低 Φ_f 值)且不具有清晰边界。看来,这些蛋白质的折叠是以穿过不同折叠核的多条平行路径而进行的,这在实验上同时观察到大量折叠中的蛋白质,并表现为弱描绘的骨架核。这种观点强调的是,在这些蛋白质中引入的一些突变可以使折叠核更准确地局部化,而且一个突变在一个地方局部化,其他突变在其他地方局部化。

课堂里的声音:根据您所说的,可以推论出,任何影响了蛋白质折叠速度的突变应该也影响其稳定性。但也有一些实验显示出相反的结果!

讲课人:我所说的这些,只适用于"全或无"型转变,在其过程中没有蛋白质链的亚稳"半折叠"状态。如果存在有这样的中间态(只有远离蛋白质的天然态和变性态的平衡点才有可能,见图 20-4 最左边部分的曲线弯曲部,而同样的自由能曲线见图 20-7),则不影响天然蛋白质稳定性的突变可以影响其亚稳的中间态,而您所说的效应(顺便说到,它是在这样的条件下被观察到的)就会成为可能。

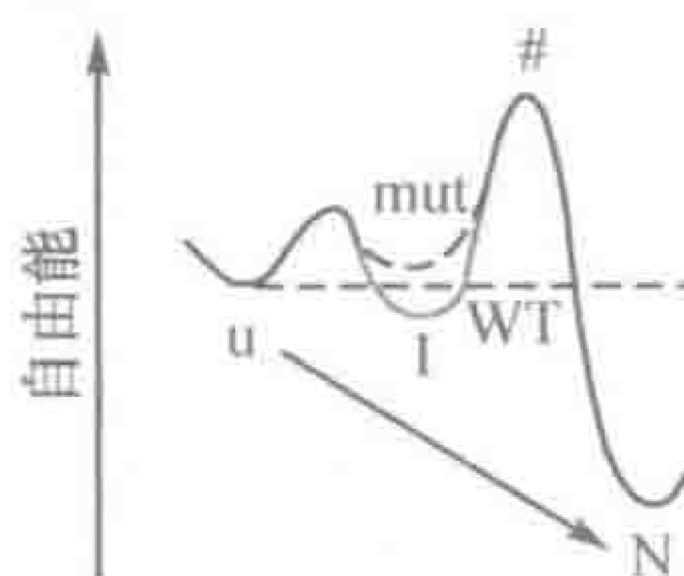


图 20-7　由解折叠态(u)通过中间体(I)到天然态(N)的自由能曲线示意图。中间体在环状型蛋白质中是稳定的(与 u 相比),在带突变的蛋白质中是不稳定的(这相应地显示在 I 区灰色实线部分和点线部分);有人认为在这里突变只影响中间体 I。在本例中,突变蛋白质的折叠进行得更快,因为该蛋白质在到达限制整个过程的速度的过渡态#之前,不需要从稳定的(与 u 相比)中间体爬上去。请注意,对于突变蛋白质,从 u 到 N 的转变是以单步骤(u→N)进行的,其中的中间体 I 与初态 u 相比是不稳定的(因而未被观察到);而对于无突变的蛋白质,从 u 到 N 的转变是两步骤的(u→I→N),其中的中间体 I 比 u 更稳定。然而状态 N 自由能的升高(N 的稳定性与 u 的稳定性更接近)会使无突变蛋白质(WT)的状态 I 失稳,而其中的转变也只沿单步骤进行,u→N。(要看到状态 I 的失稳,基本精神是,会从右边界使 WT 的曲线提高)

■　最后,我想回到图 20-4 并强调,即使对非常具体的蛋白质也不存在精确测定的"特征折叠时间"。确实,该图表明,溶菌酶在天然条件下折叠需花费约 0.1s,在特定的盐酸胍条件下(在相同的溶液温度下)达到天然形态与变性形态的平衡需要 10 000s。而在其他条件下,在较高温度但没有盐酸胍的条件下,两种形态的平衡区域对应于(图 20-2)约 10s 的折叠时间。因此,在讨论蛋白质折叠速率时,我们应该要么记得该蛋白质的所有观测到的折叠时间的范围,要么记住所有具体的实验条件,如蛋白质在细胞中折叠时的条件。

第21讲

"Levinthal 佯谬"的解决：快速折叠路径网络自动引导肽链达到稳定结构。为此必需的只是在肽链的天然折叠和其他球状折叠之间存在明显的能隙。对几种蛋白质[塞尔平(serpins)和普利昂(prion)]中形成稳定结构异常慢速的讨论。蛋白质"能景"的概念。蛋白质结构：自组织物理学和自组织链的自然选择。

我们继续讨论蛋白质的自动折叠。

■ 我在前面讲过的所有实验数据，无论它们本身是多么有趣都未能回答这个问题：蛋白质究竟是**如何**在几秒钟或几分之一秒钟之内（折叠所花费的时间）在天文数字的可能结构中找到其天然结构。

而这个数字，正如我所讲过的，还在 1968 年就由 Levinthal 估算出来了，确实是极其庞大的：一条由 100 个残基组成的链有 ~10^{100} 个可能构象；假设从一个构象转变到另一个花费 10^{-13}s，则"呆板地"遍历它们需要花费 ~10^{80} 年，而 Levinthal 认为，完整地遍历构象只能是"呆板地"进行，因为甚至只要偏离 1Å 就会大大地提高肽链在整个球体中的能量，所以蛋白质只有直接落入到各个构象中才能"感觉"其稳定性。

蛋白质在无限多可能的结构中如何选择出其天然结构？Levinthal 问道，并回答说：看来，自动折叠的蛋白质需要沿着某种特定的"折叠路径"，而该路径结束时的结构就是其天然结构，不论是否存在更稳定的肽链折叠形式。换句话说，Levinthal 认为，决定蛋白质天然结构的既**不是**稳定性，也**不是**热力学，而是动力学，即它对应的不是肽链自由能的全局最小值，而仅仅是能快速达到的最小值。

这个问题的复杂性在于，所提出的这个问题无法纯粹从实验上解决。确实是这样：假设一个蛋白质链具有另一个"天然的"、但更稳定的折叠形式，如果蛋白质本身不是处于该构象中，那又如何找到它呢？花 10^{80} 年等待结果？

另一方面，在解决各种实际问题的过程中不断地出现这个问题：是什么：动力学还是热力学决定蛋白质链的折叠？在考虑根据氨基酸序列预测蛋白质结构（需要知道预测的是什么：最稳定的还是最快折叠的结构）时出现这个问题。在考虑设计在自然界面不存在的新蛋白质（需要知道要做什么：是尽量增强所需结构的稳定性，还是铺设能最快到达该结构的路径）时也出现这个问题。

■ 然而,在“稳定结构”与“快速折叠达成的结构”之间是否真的存在矛盾?也许稳定结构**自动地**就是“快速”折叠路径的焦点,而且**自动地**具有快速折叠的性质?

■ 在着手研究这些问题之前,即考虑蛋白质折叠的**动力学**观点之前,让我们回顾蛋白质折叠在**热力学**领域的一些基本事实(在这里总是讨论相对较小的单域蛋白质,即由50~200个氨基酸残基构成的蛋白质)。这些事实将有助于我们了解,我们需要研究的折叠过程是在什么样的条件下进行的。热力学事实如下。

(1) 蛋白质的解折叠是可逆的,而且是以“全或无”型转变的形式发生。后者意味着,在蛋白质变性点以明显数量存在的只有两种蛋白质分子形态:“天然态”和“变性态”,而所有其他的形态(“半天然态”和“错误折叠态”)实际上并不存在。这种转变只在这样的肽链上才有可能发生:肽链的氨基酸序列提供了在最稳定结构与大量其他结构之间有足够大的能隙。

(2) 蛋白质(至少,对于在强烈的高浓度变性剂之下解折叠的小蛋白质)的变性态往往是无序的线团。

(3) 在正常生理条件下,蛋白质的天然态只是比变性态稍微更稳定(而在转变中点,这两种形态自然具有相等的稳定性)。因此,蛋白质的天然结构的稳定得益于其低能量,即得益于在天然结构中的强烈相互作用,而解折叠结构的稳定得益于其高构象能,即得益于各种解折叠构象的巨大数量。

必要的说明:正如文献中采用的,严格来说,术语“能量”在这里表示的是**所有相互作用的自由能**,包括肽链与溶剂相互作用的自由能(例如,你们必须记住,疏水相互作用的“能量”决定于溶剂的熵,这里的术语“熵”只包含肽链的构象熵,而**不**包含溶剂的熵)。采用这样的术语的目的是,把溶剂排除在外,把注意力集中在主要问题上:蛋白质链是如何在庞大数量的可能空间结构中找到“自己的”空间结构。

因此,要解决“Levinthal 佯谬”,并证明蛋白质链最稳定的结构可以在一个合理的时间内找到,我们可以,在第一级近似下,只考虑从线团到肽链最稳定结构“全或无”转变的速度,而且我们只需要考虑如下情形就够了:最稳定的肽链折叠只比线团稍微更稳定,而其他蛋白质链形状是热力学不稳定的。在这种情形下考虑蛋白质折叠是最简单,因为这里没有稳定的折叠中间体:正如在前面的章节,尤其是第19讲中讲到的,它们只出现在天然结构远比线团更稳定的情形下。在那种情形下蛋白质的折叠能达到最高速度,但对它们的分析比较复杂。因此我们在开始时限制在天然结构与线团的平衡点附近,在该处蛋白质的折叠虽然不是最快的,却是最简单的。而且我们应该只考虑那些能在肽链最稳定结构与所有其他折叠结构的能量之间提供很大“能隙”的氨基酸序列(图21-1):正如你们记得的,异质高分子聚合物的统计物理表明,天然球体的“全或无”型相分裂需要存在有这样的能隙,能明显地把天然结构从所有其他结构中分离出来。

■ 首先,我试图证明,在这样的条件下,小蛋白质或蛋白质域的**最稳定结构**能**自动地**成为"快速"折叠路径的中心,而且因此应该在生理上合理的时间内,几秒或几分钟,完成折叠。

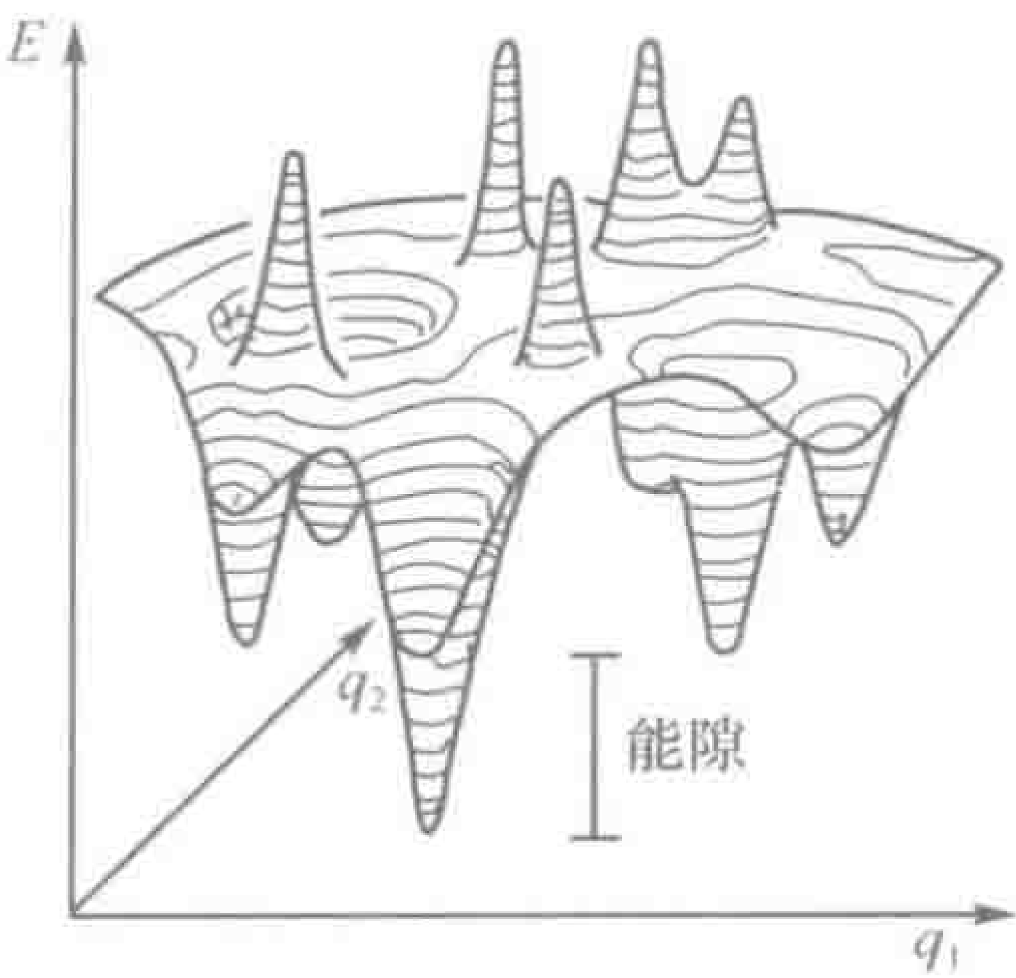


图 21-1 简略的蛋白质链能景示意图:在图中只显示了描述链构象的两个坐标(q_1和q_2),而蛋白质链的实际构象要用几百个坐标来描述。全局能量极小值和局部能量极小值之间的宽能隙是必需的,以便肽链的稳定折叠形式只有通过热力学"全或无"型转变才能破坏,这进而保证了蛋白质功能的可靠性,像电灯泡一样按"全或无"原则工作

要证明最稳定的蛋白结构应该能快速折叠,只需证明,总是至少有一条"快速"折叠路径能达到它。如果天然结构的折叠路径存在不止一个,而是很多个,则只可能是加速折叠进程。

而其他非天然结构的折叠路径(且意味着所存在的"能隙"是高能量的)我们可以完全不考虑!它们不会把我们的链从真实路径上引开。毕竟,在最稳定的链结构与线团之间的"全或无"型转变点的附近,**没有任何**"半折叠"或"错误折叠"状态可能成为陷阱:它们不可能仅仅借助微小的自身总稳定性而"吸收"折叠路径。这里有益的类比是,水通过分隔两个水池的墙壁上的裂缝渗漏:如果裂缝的"容量"很小,即它们不能吸收所有的水时,则墙上每一个新裂缝只是加速填满第二个水池,因此,在考虑水通过一个裂缝渗漏时,我们只估算它的最小填充速度。

要使路径快,则该路径上的每一步应该进行得快,这些步骤应该不会太多,而且(最主要的!)该路径不应该被某个折叠步骤上的高自由能形式的"能垒"所阻碍。

由于固定一个链节的时间很短(通过测量蛋白质链上的 α 螺旋增长速度来判断,为 1~10ns),那么如果蛋白质一个接一个地固定链节,就能够瞬时完成折叠(100 个链节的链为 100~1000ns),**如果**在这一过程中它不需要克服自由能垒的话。

■ 因此,需要回答的主要问题是:导向蛋白质链的最稳定状态的路径上的能垒高不高?

蛋白质链的折叠导致其熵降低(由于链的有序性增大)和能量减小(由于在相

邻链节之间形成接触)。熵的降低会增大链的自由能,而能量的减小会减小链的自由能。如果在折叠过程中,肽链只有在非常接近于达到自己的最终结构之后才开始出现使最终结构稳定的接触(即肽链应该在能量开始占优势**之前**降低自身的所有熵),则在折叠的第一阶段中自由能的升高将正比于链节数,即非常大,而链的折叠将非常缓慢(正如你们还记得,根据化学动力学,过程的进行时间指数性地依赖于沿其进程所达到的最大自由能升高)。这种图景(在自由能升高**之前**失去所有的熵)正是"Levinthal 佯谬",(它声称蛋白质链即使在宇宙寿命的时间内无论如何也不能找到自身的最稳定结构)的基础。

相反,如果折叠路径是这样:在进行过程中熵的降低实际上被能量的降低所补偿,则它不会被高自由能垒所阻挡,而且折叠进行得很快。恰恰是这样的图景,正如我们所确信的,确实是存在的。

在存在能垒的情形下,从经典的过渡态理论出发,过程进行的特征时间可按下式估算:

$$\text{时间} \sim \tau / \exp(-\Delta F^{\#}/RT) \qquad (21.1)$$

式中,$\tau \sim 1\text{ns}$(更准确些:$\sim 10\text{ns}$),为蛋白质折叠过程的单步特征时间(它可以根据 α 螺旋的折叠速度进行估算),T 为绝对温度,R 为气体常数,而 $\Delta F^{\#}$ 为折叠过程中的自由能与链的初始状态(线团)自由能之差的最大值的高度(或者说,过渡态的自由能)。

让我们按照如图 21-2 所示的蛋白质依序折叠的过程来分析能量 ΔE、熵 ΔS 和合成的自由能 $\Delta F = \Delta E - T\Delta S$ 的变化。在该路径中的每一步上有一个链节离开线团并占据其在最终的(最稳定的)球体结构中的位置。这个路径看上去有点像人造的(链节从哪里知道自己在最终结构中的位置呢?)。然而如果注意到,我们这种方式只是按相反方向观看蛋白质稳定结构破裂的"窗口",这种印象就会消失。

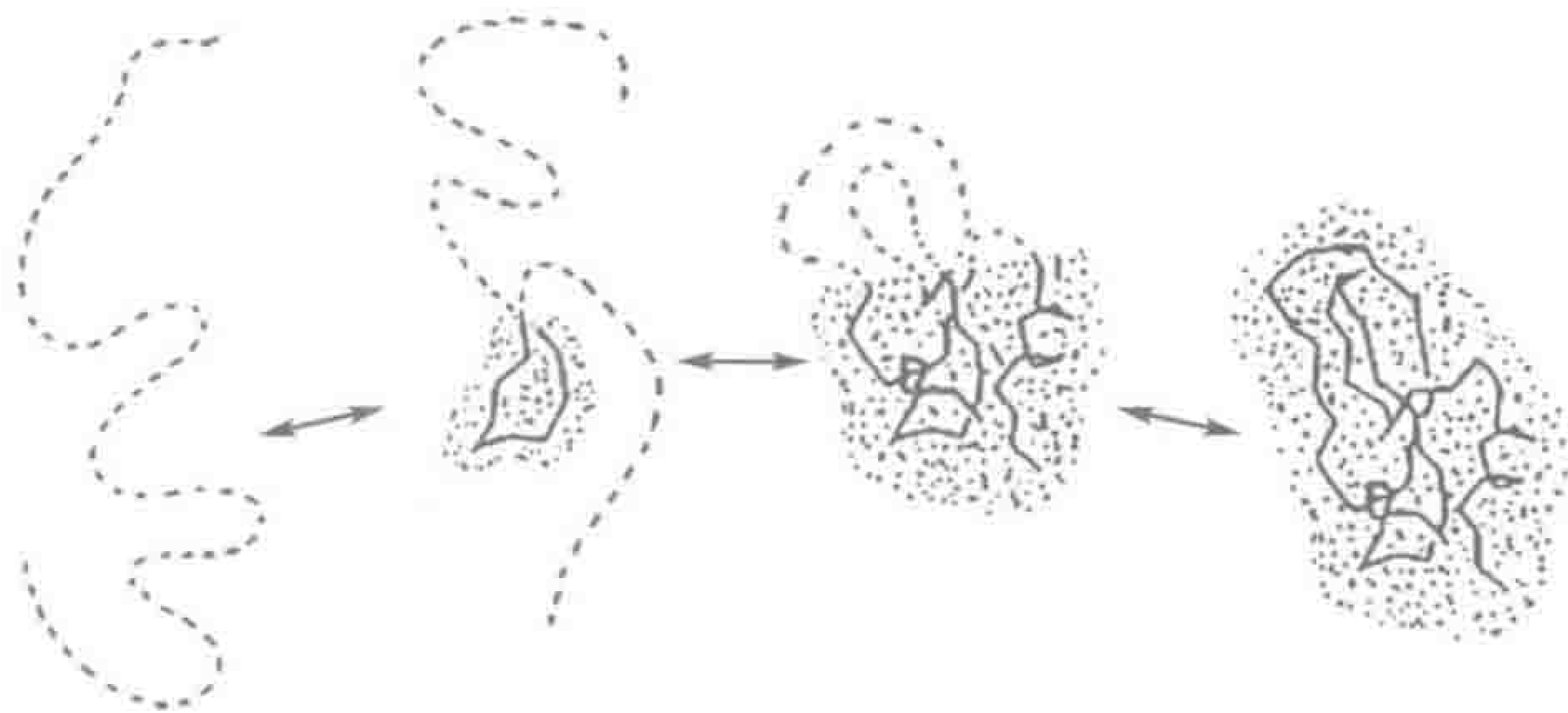


图 21-2 一条可能的蛋白质依序折叠的路径。虚线显示停留在无序状态的肽链——"线团"。以点标出的区域对应于已经获得了其最终构象的蛋白质球体部分。已形成球体的主链用粗线标出,为图片简化起见,其侧链基团未绘出。所有中间态具有较高的自由能,因此在折叠过程中没有积累,且不能直接观测

课堂里的声音:蛋白质难道不会沿一条路径折叠,却沿完全另一条路径解折叠?

讲课人:如果我们讨论的两个过程折叠和解折叠是**在完全相同条件下**(温度、pH 等)进行,则不可能。我们已经约定好在其天然态和变性态的热力学平衡点附近讨论蛋白质折叠。因此蛋白质以相同的速度折叠和解折叠——根据平衡点的定义。请记住,在 V 形图底部点附近观察到的确实就是这样。此外,折叠和解折叠应该沿相同的路径进行:如果不是这样,若蛋白质沿一条路径折叠而沿另一条路径解折叠,则会出现(在平衡条件下!)永恒的循环流动,即会出现(如果把这个永恒的流动提供给涡轮)**第二类永动机**。而正如你们知道的(应该知道!),热力学断言,第二类永动机是不可能的:这将违反熵(无序度)不减少的基本原则。

为了避免误解,说得更明确一些:**如果**折叠和解折叠是在**不同**的条件下(如在水中折叠,而在浓尿素溶液中解折叠)被研究,则这些过程可以沿不同的路径进行。但是在完全相同的条件下它们两者沿同一条轨道进行。这是普通物理学上的**精细平衡原理**①。

课堂里的声音:即使是这样……房屋爆炸的电影,甚至沿相反的方向去看,也与建造房屋的电影不相似!

讲课人:是的,但是建造和爆炸都是以花费巨大的不可逆的能量损耗进行的:燃料、炸药、人力……而蛋白质**在体外**的折叠和解折叠,都不需要任何"燃料",而且正如 V 形图所显示的,在折叠态与解折叠态的平衡点附近甚至在平衡点上,两者都可以进行。我多次强调过这一事实,因为它对理解蛋白质折叠是非常重要:一旦两个过程在相同的条件下进行,它们就会沿相同的路径进行。也就是说,如果某条路径对蛋白质解折叠是有利的,则它对蛋白质折叠也同样有利(只是要朝另一个方向进行)。

因此,在图 21-2 中所示的折叠路径是存在的,在许多其他折叠路径之中。正如我们已经讲好的,折叠速度的最低估算只要估算沿一条路径的折叠速率就够了。

在依序折叠过程中,随着最终球体中的小块的增加,在球体中使最终结构稳定下来的相互作用一个接一个地恢复起来。如果成长着的结构在整个时间内都保持比较紧凑(而正是这种含有如图 21-2 所示的紧凑的中间态的折叠路径,是我们应该感兴趣的),则这些相互作用的数量将几乎正比于被固定到球内的链节数目 n 而持续增加(而相应的,它们的能量将降低)[图 21-3(a)]。

确实,在折叠开始时能量的降低比较缓慢,因为链节黏附到较小球体的表面时生成的接触平均来说,比黏附到较大球体的表面时要少。于是,在生长球的能量中会出现非线性的表面项(即正比于 $N^{2/3}$)。因此,对能量线性降低曲线的最大偏差

---

① 精细平衡原理是化学反应动力学中的一个基本原理:在一个处于化学平衡的反应系统中,每一个元反应都与其逆向元反应具有相同的反应速率。精细平衡原理是微观可逆性原理运用于宏观的化学平衡系统时所得到的必然结果。

量约为 $N^{2/3}$，其中 N 为蛋白质链中的链节数。这个偏差显然小于链折叠过程中的能量总降低量，约为 N 大小（更准确些：约为 $N\varepsilon$，其中 ε 为残基在由线团到天然球体的转变中的能量平均变化量）。

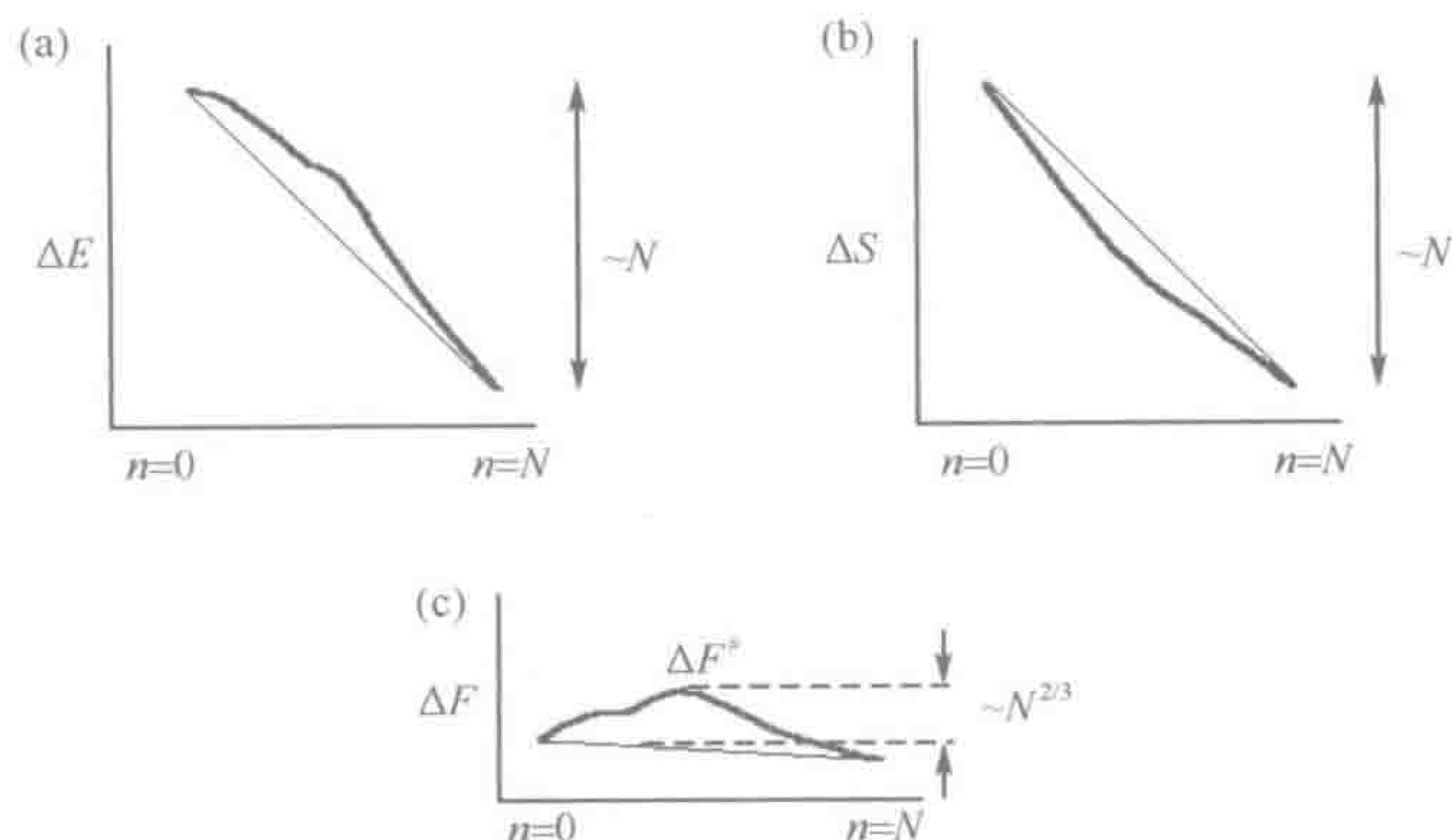


图 21-3　在最终结构与线团的热力学平衡点附近蛋白质依序折叠（$n=0$：线团；$n=N$：最终天然结构）中的能量变化 ΔE (a)、熵变化 ΔS(b) 和自由能变化 $\Delta F=\Delta E-T\Delta S$(c)。细线表示 $\Delta E(n)$、$\Delta S(n)$ 和 $\Delta F(n)$ 的线性（随 n）变化部分。$\Delta E(N)$ 和 $\Delta S(N)$ 的全部变化大约正比于 N，即链中的链节数，而粗线显示的 $\Delta E(n)$ 和 $\Delta S(n)$ 对线性（随 n）路径的最大偏差相对很小，只正比于 $N^{2/3}$。结果是 $\Delta F(n)=\Delta E(n)-T\Delta S(n)$ 对线性路径的偏差也很小（$\sim N^{2/3}$），而且在平衡点处（此时有 $\Delta F(0)=\Delta F(N)$），这就是偏差 $\Delta F^{\#}$，即自由能在折叠路径上的最大值，也只正比于 $N^{2/3}$

另外，在球体增大过程中肽链的熵降低了，大致正比于固定于球体中的链节数［图 21-3(b)］。然而，在折叠开始时，由于形成了从生长球中自由伸出的闭合环线，熵可能降低更快（图 21-4）。因此，在该生长球的熵 ΔS 中出现了非线性项（表面项），其值约为 $N^{2/3}$（与 ΔE 中的表面项相同），大大地小于熵本身的降低量，约为 N（也与能量降低相同）。

包含在熵 ΔS 和能量 ΔE 中的线性项和非线性项都必定进入到生长球的自由能 $\Delta F=\Delta E-T\Delta S$ 中。然而，当我们考虑到在给定的条件下，最终球体处于与线团的热力学平衡中（或者几乎平衡，如当它总是比线团要更稳定几个 RT），则在差值 $\Delta E-T\Delta S$ 中较大的线性项（正比于 n）**相互抵消**。事实上，在平衡点上，在线团中（即 $n=0$ 时）和最终球体中（即 $n=N$ 时）都有 $\Delta F=0$。这意味着，在依序折叠中熵的增大和能量的降低在主项上（正比于已折叠链长度 n 的线性项）相互抵消，**如果没有表面**效应，ΔF 就会在整个所讨论的折叠路径上都等于 0！

因此，自由能垒只与相对较小的非线性表面效应相关。意味着，在依序折叠的任意路径上（形如图 21-2），能垒高度**不是**正比于链节数 N（实际上，就是 Levinthal 所认为的），而只是正比于 $N^{2/3}$［图 21-3(c)］。

结果是，达到最稳定结构的时间随链节数 N 的增长不是 Levinthal 方式（即不是 10^N，且完全不是以 N 为指数），而只是 $\exp(\lambda N^{2/3})$，而且这里最重要的是 $N^{2/3}$ 值是由

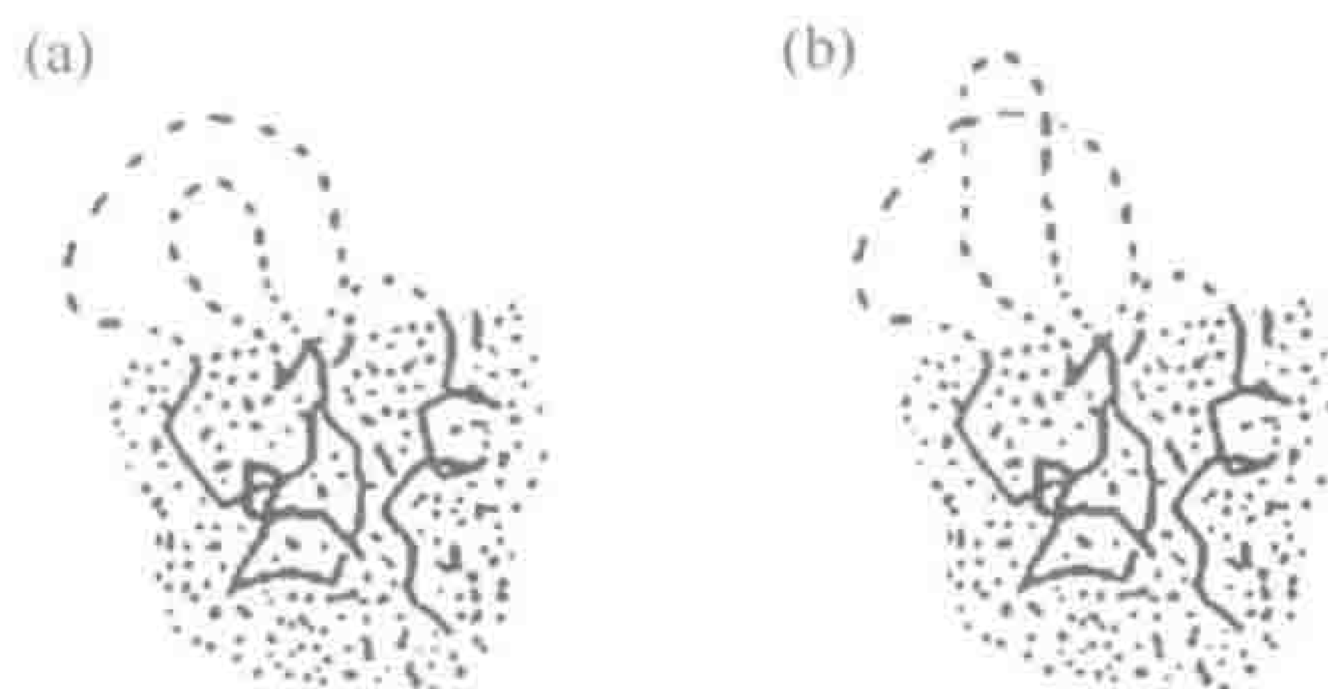


图 21-4 (a)含有从蛋白质上伸出的无序环线的紧密蛋白质折叠中间体。中间体的生长对应于蛋白质链的已固定部分(球状的)和无序部分(线团状的)的边界的移动。中间体的成功折叠要求其中的环线有正确的打结方式:含有不正确打结方式的半折叠结构(b)不能生长形成正确折叠的蛋白质,它必须先解折叠才行。不过,在由~100 个链节构成的链中可以形成 1~2 个结点,使含有不同环线打结方式的中间体的轮换不至于限制蛋白质折叠速度

于在折叠过程中肽链的天然相与解折叠相的“分离”而出现的,并且 $N^{2/3}$ 远远小于 N。同时,对系数 λ 的精确估算表明,$\lambda=1\pm0.5$,而且具体的 λ 值既依赖于肽链折叠成球的形式(是大量还是少量的不规则环线从类天然晶核上伸出),也依赖于球体中或强或弱地相互吸引的残基的分布(强烈吸引的残基是否在晶核中聚集)。

所观察到的折叠时间(在线团和天然球体这两种状态的热力学平衡点的“线团→天然球体”转变)介于从 $10\times\exp(0.5N^{2/3})$ ns 到 $10\times\exp(1.5N^{2/3})$ ns 的范围,与我们的估算相符。

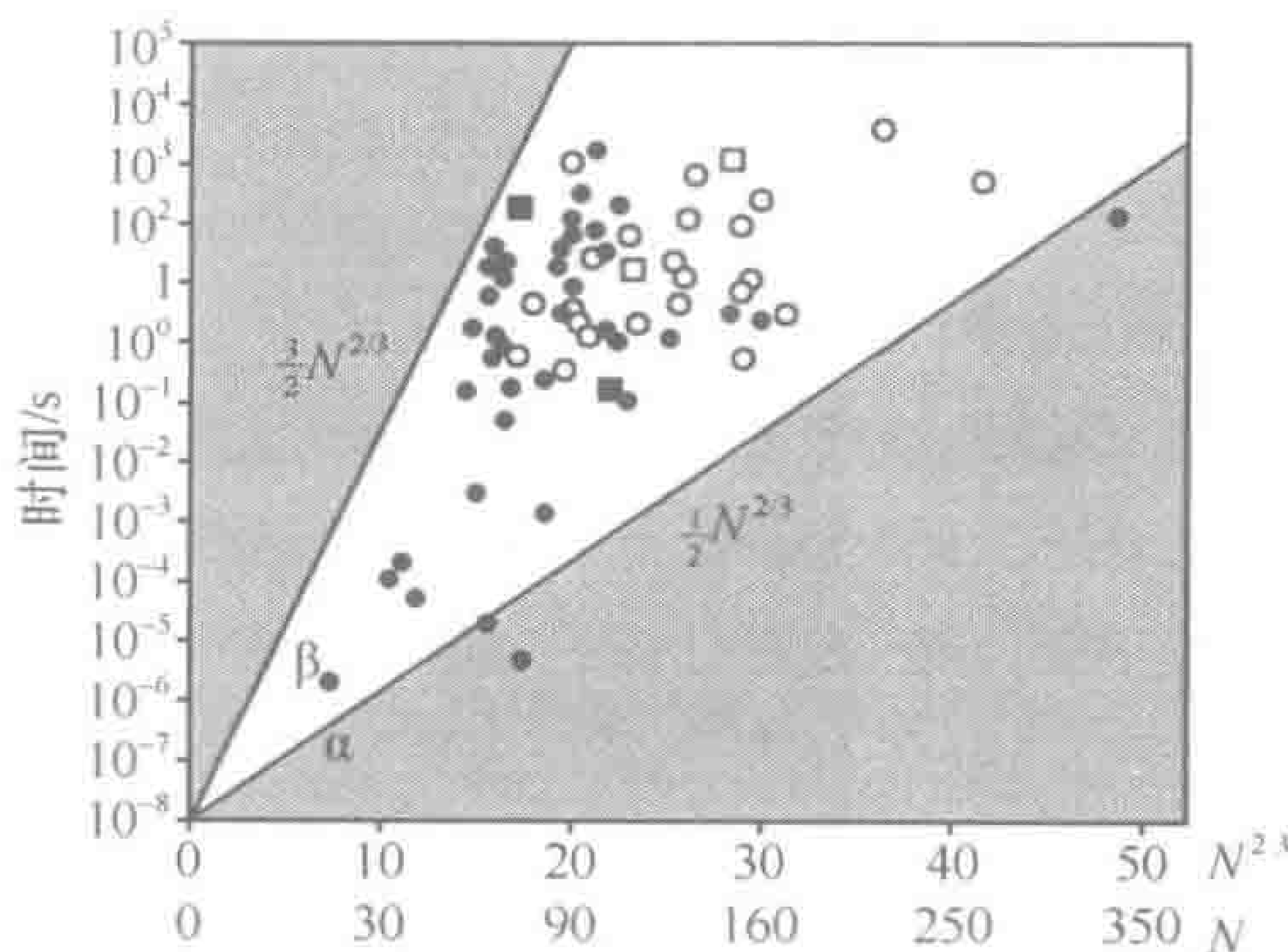


图 21-5 蛋白质折叠速度的实验数据,背景为求解得的式(21.2)的范围。数据来自线团与天然球体的热力学平衡点,即动力学 V 形图的最低处。N 为链中的氨基酸数目,圆圈为无二硫键的蛋白质,正方形为含二硫键的蛋白质,实心符号为折叠时在整个实验条件范围内无明显中间体的蛋白质,空心符号为观察到(但超出线团与天然球体的热力学平衡区域之外,接近“生理”条件)有中间体的蛋白质。符号 α 和 β 分别为单独的 α 螺旋折叠[Thompson P. A., Eaton W. A., Hofrichter J. 1997. Biochemistry. 36: 9200]和 β 发夹折叠[Mucoz V., Thompson P. A., Hofrichter J., et al. 1997. Nature. 390: 196]。图片取自 Galzitskaya O. V., Ivankov D. N., Finkelstein A. V. 2001. FEBS Lett. 489: 113-118(基于蛋白质专辑 Jackson S. E. 1998. Folding & Design. 3: R81-R91),并补充了最近 5 年的数据

蛋白质的最稳定结构可以在如下时间内达到：

$$时间 \sim \exp[(1\pm0.5)N^{2/3}]\times10 \quad (\text{ns}) \tag{21.2}$$

其原因在于，在依序折叠的过程中，熵降低立即被新出现的相互作用的能量所补偿（精确到表面效应）。

请注意，在我们考虑的路径中，蛋白质结构的球形晶核在折叠过程中**没有**被重新构建（这会需要大量时间），所有的重新排列只发生在松散的线团中（因而很快）。

无论如何，所得到的估算式(21.2)和图 21-5 表明，由 80 个或更少的氨基酸残基构成的肽链总是应该在几分钟（或更短时间）内达到自身的最稳定结构。由 100~200 个残基构成的肽链也差不多是这样，而由几百个残基构成的肽链完全无法在合理的时间内找到自身的最稳定的单域结构。它也解释了，为什么大蛋白质是由单独折叠的稳定的域所组成（按照"分而治之"的古老原则）。或者，正如在某些情况下所说的，由折叠子["folden"，来自单词"fold（折叠）"]组成；否则它们的折叠会极慢。

补充。式(21.2)估算的与其说是每个具体蛋白质的折叠速率，不如说是折叠时间范围。而由图 21-5 可见，同样大小的蛋白质的折叠速率有时候可能相差数百万倍。

Plaxco 和 Baker 证明，折叠时未观察到中间体的小型蛋白质的折叠速率的对数的差异，与他们引入的"接触度（contact order，CO）"关联得很好：

$$接触度 = \frac{1}{M \cdot N}\sum_{\{i,j\}\in M}\Delta N_{ij}$$

式中，M 为蛋白质内原子接触数目；ΔN_{ij}为发生接触（距离≤6Å）的原子 i 和 j 之间的肽链残基数；而 N 为肽链中的氨基酸残基数目。接触度反映了蛋白质的构造：沿链相近的残基发生接触很多的蛋白质（主要是 α 型蛋白质）这个值较小，而主要是沿链距离远的残基发生接触多的蛋白质则很大。

后来证明了，虽然接触度（只考虑了蛋白质链的折叠形态，而未考虑其大小）很好地适用于小蛋白质，但完全不适用于大蛋白质和很小的蛋白质。因此很自然地把接触度和考虑了蛋白质大小的因子 $N^{2/3}$进行"杂交"，特别是由于接触度在含义上与式(21.2)中我们估算的系数上下界 1±0.5 很接近：接触度与天然蛋白质球中的蛋白质链形成的环线有关联，而"1±0.5"与从"半折叠"球体中伸出的环线有关联。后来，我们和 Baker 两个研究组的合作研究结果表明，**所有**蛋白质的折叠速率都与接触度×N 值相关联，该值随蛋白质链长度 N 而增大，平均上为 $N^{2/3\pm0.3}$。这样就得到了一个有助于较好地根据蛋白质的空间结构预测其折叠速率的理论。

更进一步：还发现，折叠速度还可以仅仅与链中氨基酸残基数减去 α 螺旋片段中的氢键总数（根据 α 螺旋可以作为自动折叠的模块的假说：确实，它们在某些快速折叠的蛋白质的解折叠链中非常稳定）更好地相关联。而由于螺旋片段可以相当好地根据蛋白质一级结构进行预测（如 D. Jones 的 PsiPred 程序），现在可以**仅仅**根据蛋白质（和多肽）的氨基酸序列来预测（精度可达数量级）它们的折叠速率。

■ 附注。

(1) 求出折叠路径上的最大自由能之后,我们就可以进一步找到它在路径上所对应的点,即估算蛋白质结构的折叠核大小。计算表明,这个核相当大,它涵盖了蛋白质球体的大约一半。对此实验也证实:蛋白质的折叠时间和解折叠时间随变性剂浓度的改变速度(图 21-6)大致相同(但符号不同!),这表明在过渡态,肽链的溶剂可及表面大致介于该链在天然态和解折叠态的溶剂可及表面的中间。这么大的核尺寸表明,蛋白质不能沿很多的平行路径[具有根本不同的过渡态(和折叠核!)的路径]进行折叠:如果大紧密核有很多,它们必定强烈地重叠,所以相互独立的折叠路径数目不应该有很多。因此,我们只考虑可能折叠路径中的一条,没有犯大的错误。

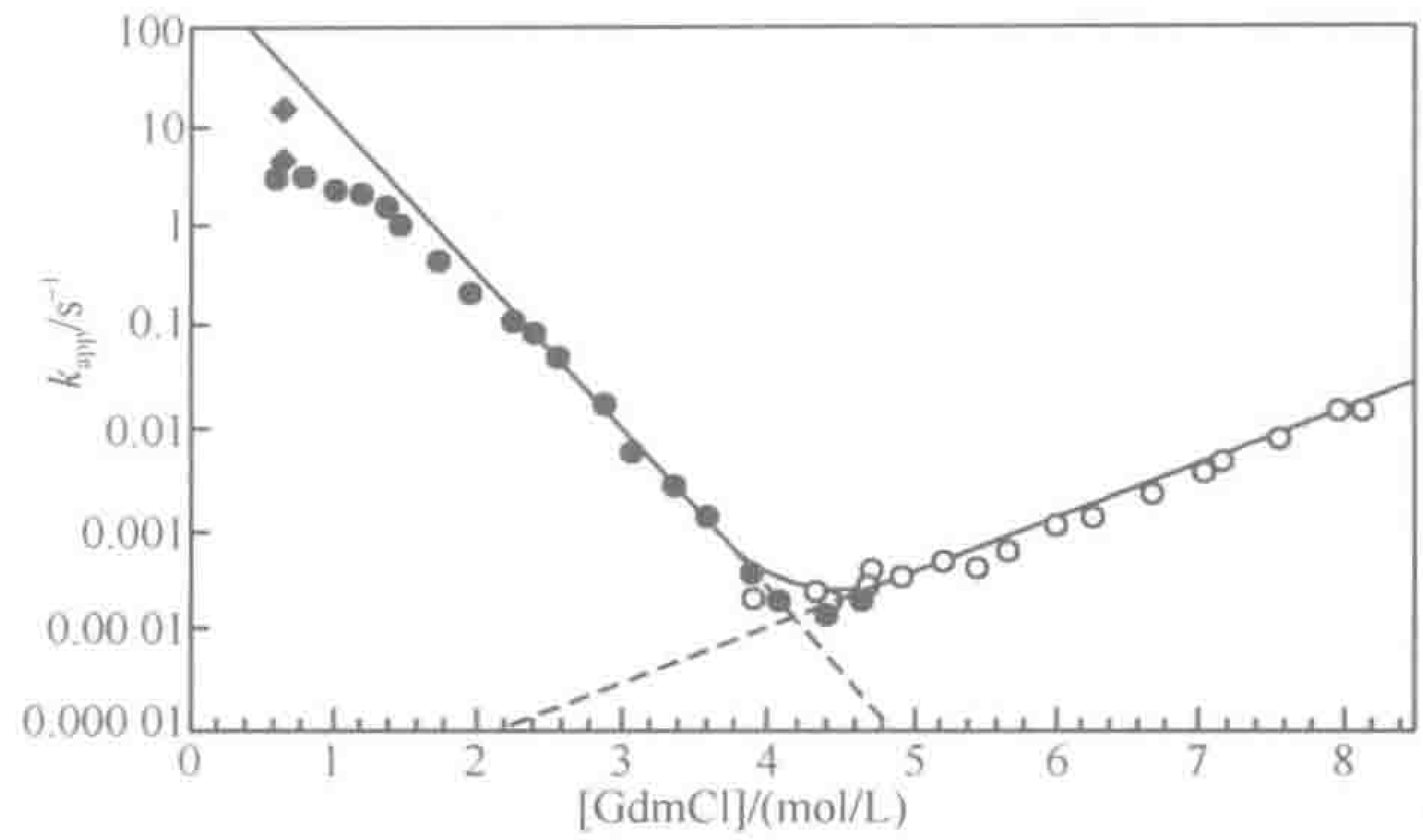


图 21-6 蛋白质(鸡卵清溶菌酶)的变性和复性与盐酸胍(GdmCl)浓度的关系曲线。实验数据点显示了邻近蛋白质天然形态和解折叠形态的平衡时的表观速率 $k_{app}=k_{u\to N}+k_{N\to u}$。实心圆点是通过把初始时含有放有变性蛋白质的 GdmCl 溶液进行稀释(即蛋白质的完全或部分复性)获得的,这种情况下有 $k_{app}\approx k_{u\to N}$。空心圆点是通过向初始时的天然蛋白质中添加 GdmCl,即蛋白质的完全或部分变性获得的,这种情况下有 $k_{app}\approx k_{N\to u}$。虚线表示 $k_{u\to N}$ 和 $k_{N\to u}$ 值在 V 形图折点处的延长线。这两条实验点连线的在图中下方部分(≈4.5mol/L GdmCl)的弯曲对应于蛋白质的天然形态与解折叠形态的热力学平衡区域(此处有 $k_{u\to N}\approx k_{N\to u}\approx k_{app}/2$)。在这些条件下复性时间为 ~$10^4$s,而在"几乎纯净"的水中,即 0.6mol/L 的 GdmCl 溶液中(曲线左上方部分),它只是约为 0.1s。图片取自 Kiefhaber T. 1995. Proc Natl Acad Sci USA. 92:9029-9033,略有修改(添加了延长虚线)

(2) 确实,我们的估算式(21.2)适用于线团与蛋白质天然结构的热力学平衡点处,在该处实验上观察到的该结构的折叠时间最大,它可能比在天然条件下的折叠时间超过几个数量级(图 21-6)。因此估算式(21.2)给出的与其说是定量的特征,不如说是原则性的特征:它解决了 Levinthal 佯谬,并解释了为什么蛋白质链找到其最稳定的结构不需要天文数字大的时间,而是在生物学上合理的时间内。

(3) 对于由几千个链节构成的长链,对各种不同的打结中间体的"准 Levinthal"轮换可能成为限制因素,因为如果不破坏中间体的球体部分,就不能解开环线中的打结[对比图 21-4(a)和图 21-4(b)]。然而,那么大的域就不能如此

［按照式(21.2)］折叠,而在由其他数百个链节组成的域中不会有很多结(计算机实验表明,形成一个结需要上百个链节),因此,对不同打结构象的轮换会使达到域的稳定结构的时间增大到好多倍。

■　到此为止我们讨论了在邻近相变点处的蛋白质链折叠,在该处只有一个(“天然的”)结构在稳定性上与线团有可比性,而所有其他的结构,甚至加在一起,都是不稳定的［图21-7(a)］。

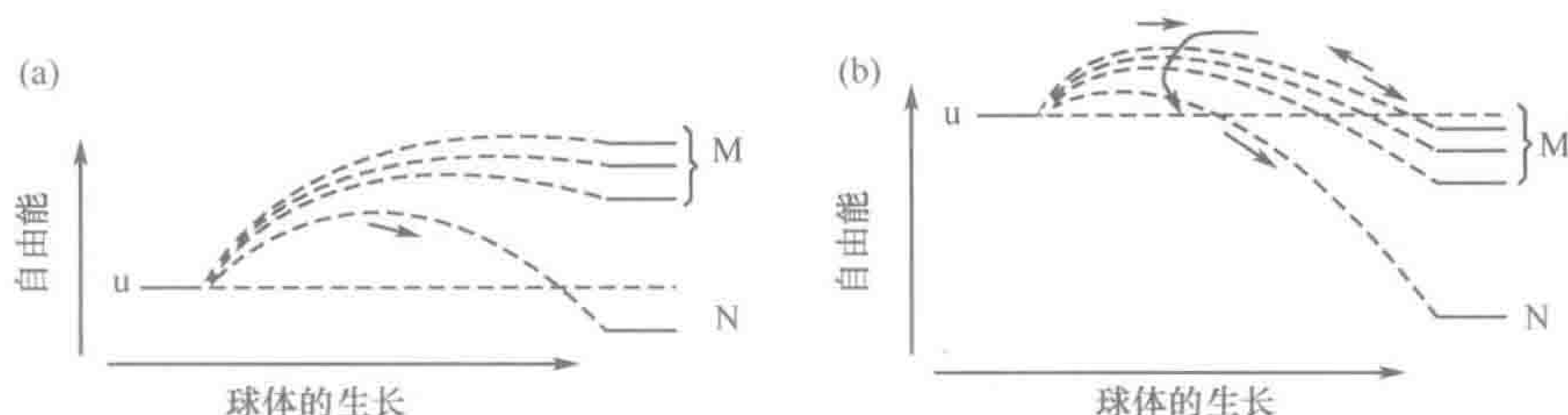


图21-7　蛋白质在不同条件下的折叠。细线表示自由能水平:线团u,肽链最稳定的(“天然”)折叠N,以及其他(与最稳定结构竞争的“错误折叠的”)球体结构M。虚线表示在折叠过程中的自由能变化。其相对较小的表面不平度(图21-3中的ΔF)在此图中未绘出。这些虚线的最高点对应于在相应的折叠路径中的过渡态自由能。折叠过程的主要情况是:①肽链最稳定的结构N比线团u更稳定,而所有与其竞争的结构M加起来也不如线团那么稳定。会发生快速且无误的最稳定结构的折叠;分子实际上不去探测“错误”结构M的折叠路径,因为在这些路径上自由能只会增大。②除了结构N之外,许多折叠结构M比线团更稳定。会发生许多“错误”结构的快速折叠,随后就出现很慢的部分解折叠,并转变为天然结构N。箭头表示蛋白质稳定空间结构的自动折叠过程的主要轨道

在远离相变点的时候,肽链自动折叠的速度会发生什么变化?当然,我们感兴趣的是天然结构的稳定性增大的情形(如果其稳定性降低,则该结构就不会形成)。

天然球体的稳定性增大时(如在温度降低或向变性剂溶液中加水时),因为其折叠中间体的自由能降低,而与天然结构相竞争的链折叠态仍然是(由于存在“能隙”)不稳定的,所以最初会导致折叠成天然结构的速度增大［图21-7(a)］。在此情况下,如果天然态和解折叠态的自由能之差从0变化到$\Delta G<0$,由于折叠核涵盖蛋白质的大约一半,则过渡态和解折叠态的自由能之差变化到大约$\Delta G/2$。这意味着它的折叠大约加速到$\exp(-\Delta G/2RT)$倍。于是,N个链节构成的蛋白质链的到达自身最稳定结构(其自由能与解折叠态之差为$\Delta G\leqslant 0$)的时间可由下式估算:

$$\text{时间}\sim\exp[(1\pm 0.5)N^{2/3}+\Delta G/2RT]\times 10\quad(\text{ns})\tag{21.3}$$

确实,这种加速大致上被观察到了(图21-6左边上升支),不过只到一定的限度(图21-6曲线左支上部的平台)。看来,最大折叠速率应该是在与天然结构竞争的肽链“错误”紧密折叠态(而在温度和变性剂稀释时它们的稳定性同样增大)的稳定性与线团相近时达到的。当球体的稳定性进一步增大时会产生大量“错误”

结构的快速折叠,然后它们慢速地部分解折叠并转变为天然球体[图 21-7(b)]

这一切都与结晶非常相似。在凝固温度下可以得到很好的大单晶,但非常非常缓慢。温度稍微下降能加速这一进程。温度下降很多会导致出现许多小晶体,而不是一个很好的大单晶。

我想提及一下,蛋白质自动折叠与一级相变在热力学上的相似性在实验上已经由 P. L. Privalov 的工作所证实,而这两者在动力学方面的相似性则由 E. I. Shakhnovich 和 A. M. Gutin 的计算机实验,以及 A. Fersht 及其同事的实验工作所证实。

■ Shakhnovich 及同事在简单的蛋白质链计算机模型上研究了最快折叠的区域。该工作就像模拟蛋白质在"天然"条件(或者甚至更有利于折叠的条件)下的折叠。他们证明,特征最快折叠时间随链长的增长比根据式(21.2)在天然蛋白质与解折叠蛋白质的热力学平衡点上估算得的折叠时间增长要慢得多。特别是,计算机实验表明,最快折叠时间随链节数 N 的增长形式不是在平衡点应该有的 $\exp(N^{2/3})$,而是慢得多:对"随机"链是 N^6,而对精心挑选以能最快折叠(并因此而在链的最稳定结构和其他紧密折叠态之间有很大的能隙)的链是 N^4。

所有这些都再次强调在上一讲中所说过的蛋白质折叠时间强烈依赖于实验条件,而且特别依赖于蛋白质天然结构的稳定性和与之竞争的结构的稳定性。

■ 这里你们自然会产生一个问题:"如果肽链具有不止一个结构以很大的能隙与其他结构相隔开,而是,例如有两个,那会怎么样呢?"

回答是:"如果这两个结构都如线团那么稳定,将会首先折叠到折叠路径更好(能垒稍低)的那一个。"然而,如果这一结构稍微不如另一个那么稳定,就会非常缓慢地(因为是折叠路径中需要亚稳结构的解折叠,图 21-7)转变到最稳定的形态。这种转变类似于晶体中的多晶态转变(回想一下"锡瘟",即白锡转变为灰锡:寒潮到来时,这种"瘟病"有时会摧毁整个仓库里的锡制纽扣)。显然,这种"多晶态的"蛋白质必定很少(理论估算表明,编码具有一个稳定折叠态的肽链的氨基酸序列是个奇迹,而编码具有两个稳定折叠态的肽链的序列,奇迹的平方)。

然而,有理由认为,某些蛋白质,其中包括塞尔平(丝氨酸蛋白酶抑制剂)和普利子蛋白(导致"疯牛病"),恰恰就有这样的行为。在普利子蛋白中,多晶形现象伴随着聚集作用(粘联),这种聚集作用能够助长(或者导向)形成"新的"β 结构形式的蛋白质。正如 K. Dobson 及其同事不久前证明的,在这种 β 结构形式下,其中的蛋白质链堆码成 β 夹层,被拼装到原纤维中,而这些原纤维被铺叠为长长的淀粉样蛋白丝(顺便说一句,类似的管状或带状的淀粉样蛋白,能形成最普通的蛋白质,如溶菌酶和肌红蛋白;在此背景上普利子蛋白非常突出,看来只是因为它的传染性)。在任何情况下,都出现 β 结构的聚集。

在聚集过程中,折叠子链的等效长度 N 增长,这(根据前面作出的估算)使得新相的出现非常缓慢。但是所出现的聚集蛋白质的新形态晶核会被拉进聚集体

中，并使所有其他蛋白质分子被引领到该新形态中（实质上与“锡瘟”中一样）。正是这种在脑细胞中的普利子蛋白的聚集，最终导致在“疯牛病”、瘙痒症，以及与之同源的有巨大潜伏期的疾病中的致命结果。最近（2010 年）有数据好像表明，这些疾病不是由普利子蛋白聚集的结果巨大的淀粉样蛋白丝引起的，而是在这种聚集过程中出现的活性“晶核”所引起。据说，老年消瘦也与这类有漫长潜伏期的疾病有关……而且一般情况下，趋向于在老年人中积累起来的各种淀粉样蛋白（或这些淀粉样蛋白形成过程中的中间体）现在正在被作为许多老年疾病的可能原因来考虑，尽管至今仍不清楚它们的巨大潜伏期是否确实与淀粉样蛋白在体内的缓慢形成有关。不过已经或多或少知道了，这些疾病是 V. E. Bychkova 和 O. B. Ptitsyn 很早以前就预言存在的“构象病”的一些变种。

然而，还存在一种更简单的、“和平的”并能很好地研究的同类的实验室例子：在试管中的水溶性多聚赖氨酸［poly（Lys）］，在 pH>10 和温度 20～50℃ 条件下能快速（几毫秒内）地从线团转变到 α 螺旋，然后（缓慢得多）从 α 螺旋转变为 β 结构。后一种转变伴随有聚集并花费几小时、几周或更长时间，其时间指数性地随 β 结构稳定性的降低而增长；对此我们已经在讨论 β 片的形成动力学时讲到过。

■ 最后我们再一次回头看蛋白质链的能景示意图（图 21-1），并看看最稳定的结构是如何自动地为自己开辟出能克服 Levinthal 佯谬的折叠路径的。

从每个能量极小值宽大地张开多少有点光滑的“小道”的“喇叭口”（或者，更准确地说，罗网）穿过起伏不平的（确切地说，甚至是陡峭的）地形（图 21-1）（“峭壁”对应着链块碰撞时剧烈、但作用短暂的排斥）。这些小道部分地对应着与所讨论的最小值相应的结构依序折叠的各条路径。在这些小道上发生相分离：紧密球与线团分离，由于球体含有部分天然结构，因而是相对稳定的。正是这使得小道是光滑的，因为线团相的熵损失几乎完全地（几乎，由于相边界区域有微小的能量性应力）由球态相的能量所补偿。在所有这些小道上不会发生球体已折叠部分的重建，因此在它们中没有很大的坑洼（而小坑洼不是障碍，如果温度不是很低的话）。分子沿着这些小道向能量极小值移动时，会得到能量，但同时，由于固定其折叠态而损失熵。能量极小值越深，通向它的小道就越陡峭，越容易克服沿着这些小道滑向链的能量极小值的折叠固定过程中出现的“熵阻”。由于这种阻抗是熵性质的，所以它的强度正比于温度。温度很高时，它把链从所有能量极小值中顶出，从而链完全不会折叠。温度非常低时，熵阻很小，链滑进任意最近的极小值中，很长时间才能从中移出并重新滑进其他某个地方，无论如何也不能进入全局能量最小值［图 21-7（b）］。最后，在对折叠最佳的温度下，熵阻只有在导向一个最深的最小值的路径上能被克服，蛋白质链也相当快地落到那里［图 21-7（a）］。

■ 上述的关于在折叠过程中熵被能量所补偿的观点解决了 Levinthal 佯谬，既可以应用于由线团形成蛋白质天然结构，也可以应用于其他场合：由解折叠球或由与之类

似的中间体形成天然结构。但在那些场合要做任何定量的估算都要复杂得多。对此我想强调的是,实验**没有**显示在熔球的形成过程中蛋白质自动折叠的原则性的、多级的加速。因此我就仅限于讨论直接由线团形成天然结构的更简单的情形。

课堂里的声音: Levinthal 佯谬难道不能这样简单地解决吗:分子在向能量极小值推进时,丢失熵,同时获得能量?为什么需你讲过很多次的相分离?也许不需要相分离,只需要线团简单的均匀聚集就足够了?

讲课人:只要看看这种解释如何能作为一般性的解释,就知道它的简单性是行不通的。首先,线团的均匀聚集不能提供"全或无"型转变。对此已经在第 17 讲的结尾处讲过了。因此,线团均匀聚集的假说不能解释在实验上所观察到的蛋白质在其变性态和天然态共存平衡点附近的折叠和解折叠。这意味着,它不能解决蛋白质在一般情形下的折叠问题。然而这种假说有可能应用于天然球体比变性蛋白稳定得多(数值约为蛋白质熔化的焓)的极端条件下。

课堂里的声音:蛋白质折叠问题难道不能用您在前面某讲中讲到的单步骤机制来解决吗?

讲课人:这种解释的简单性作为一般性解释也是行不通的。要想让单步骤机制帮助解决蛋白质折叠(图示见习题 19.1),就必须让折叠的每一步都是不可逆的,否则一切都会重新导向了肽链所有折叠变体的轮换……这意味着它只有在天然球体比线团稳定得多的情形下才能帮助折叠。但是它不能帮助解释在实验上观察到的蛋白质在其变性态和天然态的共存点附近的折叠。这意味着,单步骤机制也不能解释在一般情况下的蛋白质折叠问题,尽管它可以在天然球比解折叠链稳定得超过许多倍 kT 的情况下帮助解释蛋白质的折叠。

■ 所说的这些都凸显了发展关于蛋白质折叠的"新观点"的重要性,这种观点不是以离散的折叠中间体为基础,而是以用于确定折叠路径的能量表面或"能景"的性质为基础。这一过程与其他熟悉的在小分子上进行的反应有如下差别:①蛋白质结构是作为大量弱相互作用的结果而出现的;②折叠过程伴随着很大的构型熵变化;③在此过程中出现"天然"相与"变性"相的分离。

最近出现了一种观点认为,出现蛋白质折叠实验性质在表面上完全不同的场景的原因在于,在不同的条件下,各种相互作用在分子的总自由能中作出的贡献不同。多种观点的这种融合给出了一个梗概,在此框架内实验工作者和理论工作者可以一起工作,以便对折叠过程的特殊性质获得更详细的理解。

特别是,目前正在加紧研究折叠过程的路径与在细胞系统中的折叠蛋白质的状态多样性之间的关系,这导致了有可能而且有必要考虑更复杂的折叠图景,其中不仅包括分子内的相互作用,还包括分子之间的相互作用,与分子伴侣或其他蛋白质、DNA 和 RNA 形成晶核的相互作用。

■ 理解蛋白质自动折叠现象,远远超过了对研究者的才能和智力的挑战。蛋白质的折叠(或自组织)构成了物理学规律和进化选择(生物学过程的本质)的结果之间的桥梁,多年来,我们只能非常笼统地讨论这一问题,强调这样一个复杂的物理过程的生物学重要性。现在,我们开始明白,折叠是一个只发生在已经恰恰为了保证这一过程能有效进行而被按照生物学选择出来(请回顾"能隙")的系统中的物理过程。

蛋白质物理学建立在如下两个基本实验事实的基础上:①蛋白链(至少不太长,不会逆生物合成而被修改)在合适的周围环境中能够自发地形成其三维天然结构;②蛋白质的天然态与其解折叠态是通过"全或无"的热力学转变隔开的。尤其是后一实验事实保证了蛋白质功能的可靠性和部分折叠并因此趋于分子聚集的蛋白质最小构成。

在这几讲中所讲到的一切都导向一个观点:看来,生物进化只挑选出那些能够折叠成为很好的特定三维天然结构的序列。事实上,正是它确定了蛋白质行使其功能的稳定性。而对此需要那些其中一个结构比备择折叠结构的能量低得多的序列。而正是这样的序列的肽链在折叠和解折叠过程中能够进行"全或无"型转变。看来,天然结构的稳定性的增强与肽链紧凑拼装成球有关(虽然目前还不完全清楚,紧凑拼装成球的能力对蛋白质链提出了哪些具体的要求)。值得注意的是,按照这个征兆"中选"的链既满足能折叠成稳定的空间结构,同时也满足快速折叠的要求。看来,后一种能力能够帮助尚未完整折叠的多肽链避开分子间聚集的竞争性过程。稳定结构能快速折叠的能力解决了很早以前就注意到的按动力学(根据实现方法)和按热力学(根据结果)选择天然结构的矛盾。

球状蛋白的序列看上去相当"随机",而它们的二级和三级结构,正如我们已经看到的,具有很多的对于随机高分子共聚物的所有多少有点稳定的折叠结构很典型的特点。但是,蛋白质多肽链能够快速且可以重现地折叠为特定的空间结构的能力,无论如何也不是随机序列的特征。"类蛋白质"的氨基酸序列的基本特点,即决定它们的所有物理性质的特点,是这些序列的天然结构的稳定性升高了,即在天然结构的能量与错误折叠的球状结构的最小能量之间存在有很大的能隙。在实际蛋白质中这种能隙的大小尚未确定;然而已有的估算表明,对于单个域(由100~200个氨基酸残基组成)的"类蛋白质"特性只要相对较小的能隙就足够了:10~20kcal/mol。而且已经相当清楚,天然蛋白质结构的稳定性所必需的升高可以从非常小的、随机形成的能隙开始,并通过对稳定的(因此能稳定地工作)蛋白质有利的点突变渐进性进化选择来获得。

■ 蛋白质折叠可能是只能发生在经进化选择以允许它进行的系统中的最简单的生物形态发生(morphogenesis)的例子。然而,蛋白质折叠所具有的所有物理特性,对于遵循统计物理规律的复杂过程是很典型的。特别是,球状蛋白的折叠包含有成核作用,它通常对"全或无"型热力学转变,即一级相变,是典型的。蛋白质折叠的路径,看来不是唯一的(而且在球状蛋白中的折叠核的位置也不是唯一的)。有时候在突变蛋白质上观察到的折叠核和折叠中间体的变化(在保持蛋白质天然结

构的情况下)显示,与蛋白质的天然空间结构相比,自动折叠的路径处于不那么严格的选择控制之下:毕竟不同的路径可能会导致同一个目标,虽然速度不同。[然而,低速折叠时可能隐含着危险:K. Dobson 和 L. Morozova-Roshe 证明,“未折叠好的”和“熔化的”蛋白质倾向于聚集,有时(在没有分子伴侣的保护下!)只需几毫秒就能聚集,并与蛋白质的“正确”折叠相竞争,或倾向于形成淀粉样纤维,所有这些可能导致我们讲到过的“构象病”]蛋白质折叠是作为在其解折叠链结构中的统计涨落结果而出现,导致(在“生理”温度,pH 等条件下)肽链转变为天然态,即具有最低能量的结构。通过这种方式,从蛋白质解折叠链或部分折叠链的统计系统中出现唯一的、明确固定的蛋白质结构。

■ 上面所说的一切,再次强调了蛋白质的结构、自动折叠的能力和稳定性之间的联系,以及进化选择在形成和维持蛋白质的这些性质中的作用。

很容易想象(但很难证明),一旦抓住某个由 DNA“随机”片段编码的多肽的一个能量偶然略低的结构(而且该结构能够,虽然很糟糕,但终究做某些对细胞有好处的事),通过突变并选择越来越好折叠的链,进化压力能导致产生具有这些罕见性质(自发地自动折叠)的序列,并促使这种“原蛋白”发展出特异性功能(而且通过在由各种分子组成的复杂和拥挤的细胞环境中对反馈控制敏感的方式而发展)。

然而,这种从随机序列起源产生新蛋白质的方式,看来,并没有在所观察到的(即非最早的)生物进化阶段起到任何明显的作用。现在所有的“新”蛋白质都有其“祖先”。正如已经讲过的,人们认为,现在出现新蛋白质的基础在于两种机制,第一种是“老”蛋白质的基因复制,它让一部分基因实际上不变,来维持“老”功能,而让另一部分基因悄悄地变异,朝着“变化的”新功能漂移(在自然选择的作用下),第二种是多个分立的蛋白质合并成能行使更复杂的反应并经受更精细的(由于各域的物理相互作用)调节的多域蛋白质。请注意,多域结构对多细胞高等生物(与细菌和单细胞真核生物相比)更为典型。

然而,通过基因复制和重组以及对它们在生物体中的效用进行自然选择而产生新蛋白质的途径,并不会废止某些(“无缺陷的”)蛋白质结构,由于其稳定性(也意味着自动折叠的能力)兼容更多的编码它们的不同序列,因而允许比其他(“有缺陷的”)蛋白质结构进行更广阔得多的进化和选择。当然,自然选择可能会促使出现甚至概率最小的蛋白质结构,如果恰恰只有这种结构对蛋白质的某种有益的(对生物体来说)功能是必需的。毕竟自然选择促成了出现这样小概率的结构,如眼睛和大脑……然而,正如我们将看到的,蛋白质的功能,在任何情况下(对于水溶性球状蛋白来说),通常与它们的总体结构只是很微弱地相关。因此,不必很惊讶的是(尽管很值得注意),那些蛋白质的结构通常与在挑选氨基酸随机共聚物的稳定折叠时可以预期得到的结构相类似,即“多数原则”看来也在蛋白质结构的生物进化现代阶段起作用。

蛋白质结构的预测与设计

第 22 讲

根据氨基酸序列预测蛋白质结构的需求。根据序列同源性对蛋白质结构和功能的“识别”。蛋白质家族的一级结构谱(profile)。蛋白质结构的关键区域和功能位点。蛋白质链的稳定空间结构的释出。蛋白质结构的“模板”。我们总是被迫根据作用于链中的部分相互作用判断预测得的蛋白质结构。结果:概率性预测。使多肽二级结构稳定和破坏的相互作用。非球状多肽的二级结构的计算。蛋白质二级结构的预测。

自从人们认识到蛋白质链的氨基酸序列决定其空间结构之后,就产生了根据蛋白质链的氨基酸序列来预测其结构的问题。

产生蛋白质结构预测的需求不仅只是纯智力上的兴趣问题“这能不能做到?”,而且是由于在实验上确定蛋白质的空间结构比确定其氨基酸序列难得多。而新获得的序列“流”在不断地增长:在实现解析各种不同的生物体(从细菌到人体)完整基因组的计划的过程中,每天都会产生上百个新序列。理解蛋白质的活动机制,搜集人工抑制剂或激动剂,甚至常常只是想确定它在细胞中起什么作用,都迫切地需要知道它的空间结构……

而且(当然!),回想起预测出 DNA 的双螺旋结构对理解整个基因的机理起着多么决定性的作用,对蛋白质空间结构的预测就更加热切了。

■　现在已经知道约十万个蛋白质序列。但“总共”只有其中几千个的空间结构已被测定了[X 射线衍射或核磁共振(NMR)]。因此最近测得的序列,许多仅仅只是从 DNA 或 RNA 中读得的,即在实验中什么也没有确定由它们表达生成的蛋白质有什么功能(而且,说到从 DNA 中读得的序列,这些蛋白质在细胞中是否表达也是问题……)。

对于这些还没有用实验(X 射线衍射或核磁共振)进行研究的序列(我将称它们为“新”序列)的三维结构,我们能说些什么?

■　首先出现的想法是,基于**亲源相似性**来预测“新”序列的三维结构,或者说,基于

它们的一级结构与某个空间结构已被解析出来的“旧”序列之间的所谓“同源性”。

实验表明,甚至并不很强的序列相似性也足够有非常好的空间结构相似性(参见彩色立体图对 D-19):常常有人说,空间结构比氨基酸序列更保守(或者说,根据蛋白质的一级结构确定其空间结构的编码是简并的)。

确认序列的同源性确实是一种非常有效的阐明结构和功能的相似性的方法(而且不仅是蛋白质,对 DNA 和 RNA 片段也是如此,但我将只讲蛋白质)。

也许,值得提及的是,术语“序列的同源性”的意义并不完全单一。举个例子。假设在基因中的蛋白质起始编码之后插入一个碱基。旧 DNA 和新 DNA(以及 RNA)几乎是相同的,它们会被识别为是“同源的”(应该说,在严格意义上:它们有相同的起源)。但同时,由于阅读框的移位,新、旧氨基酸序列是完全不同的,它们之间看不出任何“同源性”(严格地说,这是不对的:它们的起源是一致的;但这个结论在含义上是正确的,因为旧蛋白质和新蛋白质的空间结构完全不相似)。

下面我将在狭义上,即氨基酸序列的相似性使用术语“同源性”,因为我们的目标是寻找空间结构的相似性。

对于足够相近的序列(图 22-1),确认它们的同源性甚至空间结构的相似性都不困难。

```
                        1         10        20
人,黑猩猩               GDVEKGKKIFIMKCSQCHTV...
猪,牛,羊               GDVEKGKKIFVQKCAQCHTV...
家鸡,火鸡               GDIVEKGKKIVQKCSQCHTV..
狗鲨                     GDVEKGKKVFVQKCAQCHTV..
锥蝇                 GVPAGDVEKGKKIFVQRCAQCHTV...
铁锈菌               GFEDGDAKKGARIFKTRCAQCHTL..
花椰菜          ASFDEAPPGNSKAGEKIFKTKCAQCHTV...
```

图 22-1 各种真核生物线粒体和叶绿体的细胞色素 c 的 N 端片段的同源氨基酸序列。粗体字母标出了与人类蛋白质中相同的残基,下划线表示与其相似的。氨基酸序列的比对取自[6]。有如此高同源性的蛋白质具有非常相似的空间结构

当家族中的蛋白质显著变动时,就比较困难了(图 22-2)。

```
E.coli  MLKQVEIFTDGSCLGN---PGPGGYGAILRYRGREKTFSAGYTRTTNNRMELMAAIVALEAL
Yeast   YNKSMNYVCDGSSFGNGTSSSRAGYGAYFEGAPEENISPLLSGAQTNNRAEIEAVSELAKKI
MMLV    PDADHTWYTDGSSLLQ---EGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALK-M
RSV     PVPGPTVFTDASSSTH---KGVVVWREGPRW--EIKEIADLGA--SVQQLEAREAVAMALL-L
HIV     IVGAETFYVDGAANRE---TKLGKAGYVTN-KGRQKVVPLTNT--TNQKTELQAIYLALQD-
        |---β1---|          |---β2---|         |---β3---||---α1---|
```

图 22-2 细菌(大肠杆菌,*E. Coli*)、真核生物(酵母,Yeast)和三种不同病毒的核糖核酸酶 H 的 N 端片段的氨基酸序列。在本例中序列的比对是手工完成的,没有使用计算机,以便排除在它们的 α 和 β 结构片段内的空位(见“- - -”)。在这 5 个序列中的 3 个或更多序列中相同的残基用粗体字母标示。黑点(序列上方)标示了活性中心的残基,空心圈和菱形标示的是包含在该蛋白质的两个疏水核中的残基。下方(在序列比对的下方两行)标出了在 RSV 和 HIV 序列中相同的(=)和相似的(:)残基,还指示出了这些蛋白质的二级结构。图片取自文献[7],略有改动

在这种情况下，计算机来帮忙了。已经开发出了多种寻找同源性的程序，可以通过国际互联网使用它们。下面只列出这些程序中最流行的几个：BLAST，PSI-BLAST，HMMer 以及 Smith-Waterman。它们都能进行序列比对(alignment)，以努力找到它们中最大的相似性。为了提高相似性，常需引入序列“空位”(见图 22-2 中的“-”符号)。然后程序估算所比对的序列的相似性，并报告：①它们是否同源(即它们是否在遗传亲缘上相关)；②这些序列的最佳比对如何。确实，在它们的相似性不高时也能确定序列的亲缘性，但是(注意！)这时常常并不能达到建立它们的正确的(根据它们的空间结构对照而得的)比对；对此我还会回头讲到。

■ 对于残基匹配、相似、不匹配、空位起始和空位延长一个残基等性质的分值，不同程序的计算方式不同。匹配、相似和不匹配的值是从明显同源序列的氨基酸替换的统计学中得到的，而空位罚分是作者通过能满意地挑出那些已知相似的蛋白质：其相似性已从其他数据得到，(并能剔出公认的非同源蛋白)而选取的。优化的参数然后被“封装”在程序中。使用程序时，一般人(“用户”)通常只是模糊地知道，根据该程序，什么是“好的”，什么是“差的”，而且被简单地告知：“已求得，序列的同源性为 25%”，即比对的残基有 25% 彼此重合。

这就产生了一个问题：这 25% 证实了序列的相似性吗？通常来说，这与比对的长度有关(越长则越可能)，与比对中引入的空位的数量和长度有关(越少则越可能)。好的程序通常会显示，观察到的相似性可能是随机得到的概率。不过，**使用同一程序**比较公认已知的**不**相似序列，总是有益的。那样能够看出，已知的不相似蛋白质的“同源性”(图 22-3)常常达到 10%~15%，有时为 20%，间或(如果在序列中有空位)甚至 25%！

rop ——**MT**KQE**K**TALNMAR**F**IRSQTLTLLEKLNELDADE**Q**A**D**ICES**L**HDHADELYRSCLARF**G**DDGE**N**L——
mjc SGK**MT**TGIV**K**WFNADKG**F**GFITPDDGSKDVFVHFSAI**Q**N**D**GYK**SL**DEGQKVSFTIESGAK**G**PAAG**N**VTSL

图 22-3 **不**相似、**非**同源的蛋白质的氨基酸序列比对[此处是 α 螺旋形的 RNA 结合蛋白质(rop)和 β 结构型的冷休克蛋白(mjc)①]常常有 10%~15% 的匹配氨基酸残基[本例中为 69 个残基中的 10 个(见粗体字母)，即 14.5%]。比对是用 BLAST 程序进行的，不允许在序列中有空位。因此，甚至在序列中无空位的情况下，公认已知不同源的蛋白 rop 和 mjc 的比对得到了与 RSV 和 HIV 核糖核酸酶的 N 端片段的比对(图 22-2)相近似的残基重合率

这些数字随程序不同而不同。不过，积累的实验表明，当(公认的)“好”程序对序列进行比对时，如果超过 30%~35% 的残基匹配，则所显示的序列同源性、亲源性可以毫不犹豫地信任(预先声明：比较的序列长度应超过 50 个残基，超过 100 个更好)。确实，应该承认(参见彩色立体图对 D-20)，序列之间 30%~35% 的同源性

---

① 微生物在适应外界环境急剧降温的条件下都会发生应激反应，产生的一系列蛋白质被称为冷休克蛋白。冷休克蛋白(长度约为 70 个氨基酸)对乳酸菌适应低温环境和增强抗冻能力方面发挥着重要作用。

能可靠地(作为标准)证明它们的亲源性,它能达到使其结构的70%~80%正确地彼此重合,虽然还不能正确地预言其余20%~30%的重合。如果需要正确地摆放"新"蛋白质的主链结构的95%,只有当它与结构已知的蛋白质的同源性达到40%~50%才行。

如果所比对的序列的相似性不超过10%~15%,则通常无法求得它们的亲源性:这种相似性处于"随机噪声"的水平(然而,这并不是蛋白质不相似、不同源的证据,我还要回头讲到这个)。而"模糊地带"覆盖了从15%到25%甚至到30%:蛋白质看上去是同源的,但谁能保证呢? ……(在同源性为15%~20%时,序列比对与空间结构比对也有很大不同)

遗憾的是,正如我已经讲过,在不同的程序中(以及它们的不同操作方式下),所有这些阈值都不完全相同,而这些数字的特征"信任阈值"通常没有随程序附上(除了BLAST型的程序,列出了所获得的相似性的统计可靠性;确实,所有这些数字在原始文献中是有,但谁去读它们呢……)。因此,我建议,在信任任何这类程序之前,在你所知的长度大致相同的蛋白质(相似的,以及**不**相似的)上检查它(程序本身,以及你的操作方法),并理解"什么是好的,什么是差的"(另一种方法是:读原始文献……)。

而且,由于不同的作者对序列空位按不同方式评定"罚分",得到的相似性的所有这些可靠性和**不**可靠性的估算值随程序不同而"漂移"。如果罚分相当高以致禁止空位,则对20个字母的序列(蛋白质)的随机比对的相似性总是在5%水平。如果空位罚分设定为**零**,则允许"免费"插入任意空位,则蛋白质(总共20个字母)序列的随机比对给出的相似性在30%~35%水平(而4字母的DNA和RNA序列在65%水平)! 实验证明,当空位起始的罚分值等于二至三个其他氨基酸残基匹配的得分,而空位每延长一个残基的罚分等于该值的1/100~1/20时,能够实现从"不相似的"蛋白质序列中最优地选出"相似的"序列。

对寻找同源性的算法中的数学,我故意什么都不讲。这将会把我们带得太远了。不过,我想提到一个关键词:"动态规划"(dynamic programming)。这是应用于对一维系统(而序列正是一维系统)进行优化(尤其是从其他序列中优选一个序列)的最强大的方法的名称。

■ 如果序列的相似性低于30%,即它们处于"模糊地带"甚至更低水平,能不能辨认出它们的同源性和亲缘性? 有可能,但为此必须对**多**序列进行工作。

能用多序列进行比对,正是"最先进的"程序PSI-BLAST和HMMer有别于"简单的"、只能用单序列进行工作的BLAST程序。

用多序列进行工作的一般流程如下。首先对这两条待比较的肽链分别选出明显同源的序列(相似性不低于~40%),然后不是对单个序列进行互相比较,而是对这两个家族进行比较,特别要注意它们的最保守位置。若结果表明这两个家族相似(或者,如果某些序列同时包含在这两个家族中),则证实了(尽管是间接地)所比较序列的相似性。而且在评价相似性时,HMMer(Hidden Markov Model,隐马尔可夫模型)程序在家族中每一个序列的比对中,使用的不仅是已揭示的序列中的保守位置,而且利用了已揭示的相邻位置之间的保守关联性。通过这种方式有时能

把“识别阈值”推进到原始序列的相似性为 10%，**如果**这个 10% 的残基相似性位于最保守的位置，并由所比较序列的残基之间的相似关联而得到了提升。

图 22-2 显示，如果看整个肽链的话，人类免疫病毒（HIV）的核糖核酸酶 H 与其他核糖核酸酶 H 的相似性并不高，“在噪声水平”（因为在 60 个比对残基中它与 RSV 核糖核酸酶 H 总共只有 9 个相同：15% 匹配）。然而这一相似性被揭示出恰恰位于所有其他核糖核酸酶彼此相似的那些关键区域。这大大提高了该相似性的可靠性。而如果还考虑到，这些匹配的“关键区域”涵盖了活性中心的所有氨基酸残基，而且相似性在二级结构区段中得以保留，以及它涵盖了蛋白质疏水核大约 30%（而不是 15%）的残基，把它认定为正确的同源性识别是高度可靠的。

■ 为了识别“新”序列的同源性，还可以方便地使用“共有序列”（consensus sequences）（图 22-4），它们是从已经研究过的蛋白质家族中提取出来的，并强调了它们的最保守的性质[如果那些保守序列带有每个肽链位置的残基出现频度，则它们被称为蛋白质家族的“一级结构谱（profile）”]。

对于蛋白质链空间结构的识别，一级结构谱特别有用，因为它们不是基于蛋白质一级结构的相互叠置，而是基于空间结构叠置而建立起来的（这些一级结构谱被收集在，尤其是，由 Cyrus Chothia 研究组①研制的广为人知的数据库“Superfamily”中）。正是这些一级结构谱最精确地标出了对于维系蛋白质结构最重要的肽链区段。

```
                          M
        K                 F
        I                 K                 E
        V F V   G       V G       L         C
S D     A T R   A T     I A       E         G
K K     S S K   R S     S N       Q         A
N Q     E A N   K N T   T Q       N T     G Y
E E   S D D T   D D V   E E Q     S A     A I
A A   N N E S   Q R L   D S R A G Q       S L
P P G D P K A G E K I F K T K C A E C H T V
                              = - - = =
 -1 +1              10                    20
```

图 22-4 真核细胞线粒体细胞色素 c 的 N 端片段中不同位置的氨基酸成分。链上最重要、最保守的残基是那些被确定为单值的。下划线标出了“位点”Cys-X-X-Cys-His，该位点在绝大部分细胞色素（不只是细胞色素 c，也不只是真核细胞）中负责与血红素结合。氨基酸序列比对取自文献[6]

■ 在识别蛋白质的功能相似性中，同样也必须注意许多功能已经确认的“位点”：保障这些功能的比较短的序列（图 22-4 中在细胞色素中结合血红素的位点 Cys-X-X-Cys-His）。这些位点被收集在数据库中，而且有专门的程序（其中 PROSITE 是最流行的）进行检索。

---

① 英国剑桥大学分子生物学实验室的理论和计算生物学研究组，研究领域：结构基因组学和蛋白质结构。网址：http://www.mrc-lmb.cam.ac.uk/genomes/TCB/

■ 在根据与结构已知的蛋白质的同源性而建立“新”蛋白的结构时，应该清楚地记得，空间结构的相似性不能扩展到序列强烈发散的区域。那主要是蛋白质链的环线区域和不规则构象区域。在这些区域必须采用其他**同源建模**的方法进行构象计算，而这些方法至今也并不很成功，对此我将不去讲述。

此外，实验表明，甚至完全一致的氨基酸序列片段也不总是能保证它们的空间结构相同。例如，有时候甚至由 6~10 个残基组成的同一序列在一些蛋白质中包含在 α 螺旋中，而在另一些蛋白质中包含在 β 结构中。

■ 现在来看那些与空间结构已解析的蛋白质没有明显同源性的“新”序列的空间结构预测。

现在有大约一半的“新”序列属于这一类。因此，对许多在基因工程中获得的序列，我们无法根据它们与已知蛋白质的同源性（要么是太弱而无法检测出来，要么是缺乏同源性）来猜测出它们的空间折叠。因此就出现了不是基于同源性、而是基于物理学来解决根据氨基酸序列来预测蛋白质空间结构（甚至在将来预测其功能）的迫切需要（图 22-5）。

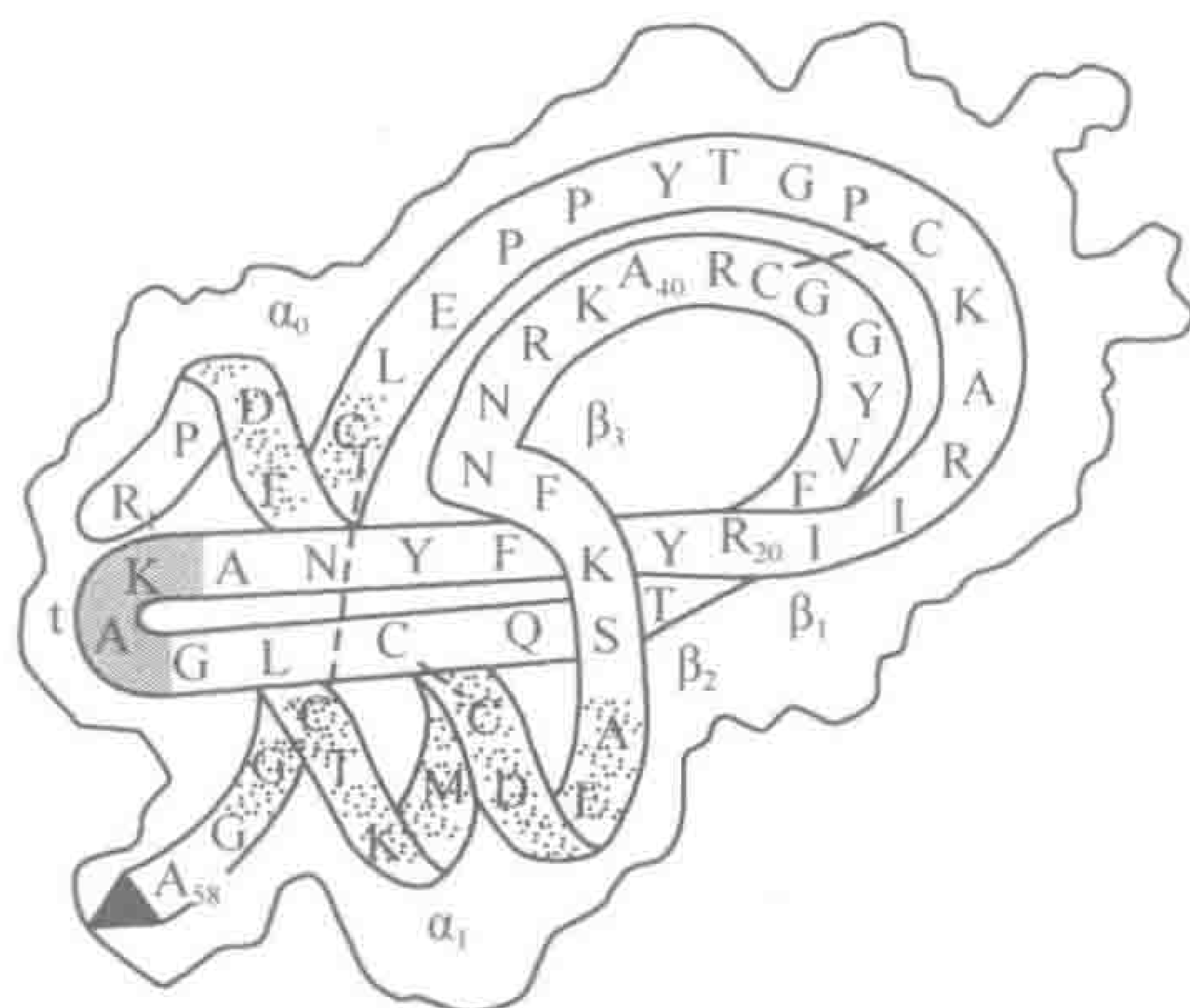


图 22-5 小蛋白质（胰腺蛋白酶抑制子）的一级结构和空间结构示意图。以分子的基本轮廓线为背景，画出了主链通路；标出了 α 螺旋、β 串、链的急剧转折(t)和半胱氨酸之间的二硫键（- - -）。因为蛋白质能自发折叠，故（**在原则上**）所有这些能够仅仅根据蛋白质的一级结构进行预测。图中没有画出侧链基团，但在原则上它们的构象也能预测

■ 直接说吧：对那些与已知结构没有同源性的蛋白质，现在没有可靠而准确的方法预测其结构。

原因（除了技术上的原因之外：甚至对于当代的计算机来说都很巨大的计算量）看来有两个：①作为蛋白质结构理论计算的基础的能量估算和其他参数的准确度有限；②在蛋白质链的一个“正确”构象和大量“错误”构象之间的能量差相对很小（因此对所有非精确计算是容易被抹去的）。问题尤其在于已讨论过的“能隙”的宽度相对很窄。如果能隙的宽度约为 10 kcal/mol，而蛋白质的熔化能量约为 100 kcal/mol，即对于可信的预测（我这里省略论证）需要能量计算值与实验值有 90% 的相符度，而现在看来是达不到的。我想特别强调后一原因，即能隙的宽度很小，这在蛋白质（以及 RNA）中的情况和在 DNA 中的情况有很大区别，这让很多从事根据氨基酸残基预测蛋白质结构的人们极为痛心。在 DNA 中，在整个双螺旋长

度上观察到有核苷的互补配对,而这保证了一条链根据自身一级结构严格与别一条链配对。它也使 DNA 的"正确"结构相对所有"错误"结构有巨大的自由能优势。核苷的这种配对的"互补编码"在 RNA 中也有,但那里互补片段比 DNA 要短得多,而且"编码"不总是能严格地遵守。而在蛋白质中**没有**残基粘连的任何单一"编码":这里既没有观察到相互接触的氨基酸残基的任何严格的成对互补,也没有由此而得到肽链"正确"折叠的巨大能量优势……

然而,现有方法能给出关于蛋白质的可能构造(或者,更经常的是,可能构造的变体)的,或者尤其是其结构元件的更为完整和可靠的信息。

在实用主义的观点上,蛋白质三维结构预测问题现在变成了下面的形式:所研究的蛋白质序列的空间结构与某种已知的蛋白质空间结构相似吗?如果**相似**,则**如何**把所研究的序列比对到该结构上?如果**不相似**,我们是否可以指出所研究序列的空间组织的某些特征细节?

我知道一些实例,包括有些来自本人的实践,当时对这些问题的回答大大促进了所研究的氨基酸序列的结构和/或功能的实验确定,帮助了制订蛋白质工程实验计划等。不过,应该说,我会有很大的兴趣阅读关于蛋白质结构预测的实际应用的好综述……

■ 现在来看蛋白质结构预测的***方法***。在讨论这些方法时,我将特别强调这些方法中的思想和它们所基于的主要物理思想。

蛋白质结构预测有两种策略。第一种策略是把蛋白质结构作为折叠的动力学过程的结果而搜寻。

第二种策略是把蛋白质结构作为对于给定肽链可能的自由能最小的结构而搜寻。

看上去,在原则上这两种策略都能得到正确的结果,因为尽管有 Levinthal 的顾虑,但蛋白质链的最稳定结构形成得相当快(这个问题已经有单独的一讲讨论,现在我不涉及它)。

然而,重要的是,在讨论蛋白质结构预测时,我们可以不考虑折叠途径而只讨论结果:蛋白质的稳定结构。而且,第一种策略,即蛋白质折叠过程的仿真模拟策略,至今没有达到系统性的进步:所有计算极其复杂,而且所有相互作用的势能都很粗略。而对蛋白质折叠的特殊原则(如分级规则)的推测(看上去很诱人,而且它们许诺能大大地简化对天然结构的搜寻)显得都不可信。因此我们在后面再讨论蛋白质折叠的模拟。

第二种策略显得更为成功。

对不同结构的稳定性(或者,等同于形成概率)的评价或是基于相互作用的势能,或是(更常见)基于不同结构元件和不同相互作用在蛋白质中的出现频率。这里没有原则性的差别:正如我们已知,蛋白质结构的准玻尔兹曼统计是由相互作用的势能所决定的。因此甚至即使它们的评价起源于蛋白质统计,而不是物理实验,

我也要将谈谈“相互作用的势能”。与水分子的相互作用通常不直接讨论,而是用由水间接参与的相互作用(如疏水相互作用)的势能来表示。

■ 让我们从根据蛋白质氨基酸序列来预测稳定的二级结构开始;α 螺旋和伸展的 β 区段是最重要的蛋白质球元件,正如你们记得的,在很多方面决定蛋白质球的总体构造。

让我们暂时忘记蛋白质链是被拼装为硬球的:在计算二级结构的同时还计算三级结构太复杂了。在预测三级结构**之前**,我们是否能根据氨基酸序列预测二级结构?看起来,是的,基本上是的,但不是绝对准确的。

首先,哪些氨基酸残基能使分立的二级结构(如 α 螺旋)稳定,哪些残基会破坏它们?

实验对这个问题给出了直接的回答。我指的是评价氨基酸残基形成 α 螺旋和 β 结构的大量工作,它们是在 Scheraga, Fasman, Baldwin, Fersht, Serrano, DeGrado, Kim 的研究组和很多其他研究组(包括在我们:O. B. Ptitsyn 和 V. E. Bychkova 的研究组里)中进行的。除此之外,对蛋白质中的 α 螺旋和 β 结构的氨基酸序列统计学也获得了丰富的信息(与物理化学实验很好地相符并对它有所补充)。图 22-6 总结了通过所有这些方法得到的评价中最重要的(即值得记忆的)结果。

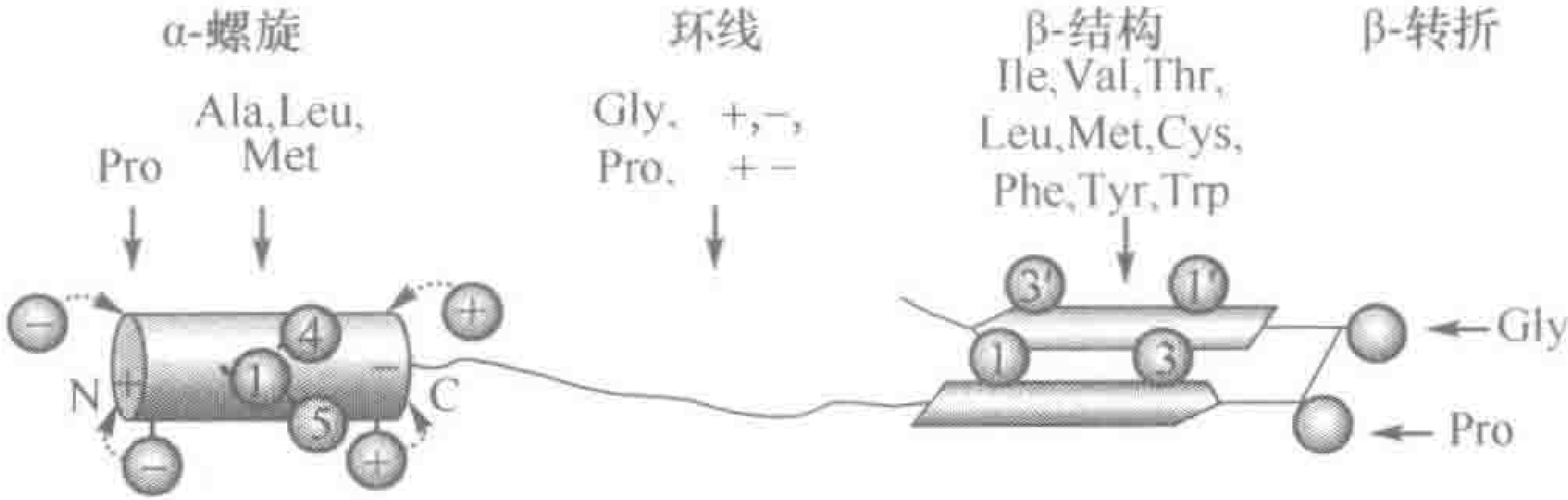


图 22-6 用于预测 α 螺旋、环线、β 结构和 β 转折的“模板”;标出了那些使它们或其局部稳定的残基“+”表示所有带正电的氨基酸,“-”表示所有带负电的,“+-”表示在侧链中有电偶极子的氨基酸。图中还显示了链上使 α 螺旋和 β 结构稳定的疏水基团的交替顺序(见编号的基团)。这样的交替方式还导致形成 α 螺旋和 β 串的疏水表面和极性表面

然而,在这里应该立即指出,在蛋白质结构中观察到的所有规律性都是概率性的**频率**特征。尽管有过许多努力,但是未能找出任何蛋白质结构的精确“编码”,即任何与核苷酸 A-T 和 G-C 配对(DNA 双螺旋的性质)相似的规则。而且,应当认识到,在 RNA 中(与 DNA 不同的是,它们的空间结构种类繁多)核苷酸配对远不是单一方式的……

绝大部分从实验和统计学得到的评价都能很容易从氨基酸的物理学和立体化学出发来理解。我们对此已经在前面的一讲中讲过了。

例如,Pro(脯氨酸)既不喜欢进入 α 螺旋(除了它的 N 端末圈),也不喜欢进入 β 结构。为什么?因为它没有 NH 基团,它也不能结合成与 α 螺旋和 β 结构相应

的氢键;而在螺旋的 N 端的 NH 基团不需要结合成氢键,因此 Pro 经常在那里出现,更何况 Pro 在主链中的两个旋转角之一(ϕ 角)已经被脯氨酸环所固定在适当的位置。根据同样的原因,Pro 也经常在 β 转折的 N 端出现。

此外,Ala(丙氨酸)能稳定 α 螺旋,而 Gly(甘氨酸)既破坏 α 螺旋,也破坏 β 结构,却有助于形成不规则区段。这种差别与什么相关?原因在于,虽然在 α 螺旋中这两个残基的可容许构象几乎是一致的,但 Gly 在线团中的构象区域,即 ϕ 角和 ψ 角的可容许区域,比 Ala 的要大得多。

同理,含 C^{β} 支链的残基(Val、Ile 和 Thr)更能稳定 β 结构,在那里它们的侧链基团有 3 个容许的旋转异构体,比在 α 螺旋或线团中要多,在那里它们只有 1 个异构体(在每个 ϕ 和 ψ 值下)。

总体上,疏水基团倾向于更经常进入 α 螺旋和 β 结构,因为在那里它们可以"免费地"(已经用氢键付过账了!)粘接成疏水簇(图 22-6),而不喜欢进入线团中,因为在线团中它们不能那样做。而有电偶极子的侧链基团,尤其是在短侧链中,更侧向于进入肽链的不规则区段,那里它们能够结合额外的不规则氢键,它们不愿意到 α 螺旋和 β 结构中去,在那里氢链的供体和受体已经被肽链内部的连接所饱和了。

氨基酸残基对二级结构的影响可以不仅予以解释,而且可以从理论上先验地估算。因为,还在得到实验测定结果之前,在 20 世纪 70 年代中期,我们(A. V. Finkelstein 和 O. B. Ptitsyn)就提出,带负电的残基应该能够稳定螺旋的 N 端,而使 C 端失稳,这是由于它能够与螺旋偶极子的局部正电荷相吸引、与螺旋偶极子的局部负电荷相排斥。而带正电的残基应该以相反方向起作用。因此每一种这样的相互作用的势能应该(根据理论估算)为 1/4kcal/mol 或 1/3kcal/mol。这已经被实验所证实。

■ 如果已知哪些氨基酸残基能稳定螺旋中间,哪些能稳定螺旋的 N 端和 C 端,我们就能得到某种像螺旋"模板"的东西。

α 螺旋的模板可以这样描述:如果蛋白质链片段的开始处富含负电基团且还有一个 Pro,中间富含 Ala、Leu 和 Met 残基而没有 Pro 和 Gly;而且 C 端包含带正电的侧链基团,则摆在我们面前的是 α 螺旋区段。除此之外,要有益于 α 螺旋的稳定性,也有益于使它包含在球体中,则特定的(图 22-6)疏水基团在肽链上的交替顺序完全是必需的:这种顺序有助于粘接这些基团,而且还能导致在螺旋上形成连续的紧密疏水表面,这对使螺旋黏附在球体上是必不可少的。侧链基团的正确交替对于形成二级结构的重要性是由 V. I. Lim 所证实的。

这就是说,"模板"定性地描述了适合于形成 α 螺旋的氨基酸序列。氨基酸序列越满足这个模板,则在肽链的该位置就越有可能是螺旋。对适合于形成 β 结构的区段也可以编制出这种描述:"模板"。对特别适合于形成 β 转折和环线的区段也可以编制出这种描述。

而且,"模板"还可以用于描述能形成更复杂结构的小段肽链,如用于描述由两个平行 β 区段和它们之间的 α 螺旋所组成的 β-α-β 超螺旋(图 22-7)。这种结构对与核苷酸相结合的域是

很典型的。在“模板”中发挥尤其重要作用的是所谓“关键位置”,它们只能由严格确定的氨基酸残基,如 Gly,所占据。只有 Gly 能处于 $\phi\approx60°$的构象中,而这种构象对所有其他残基都是禁止的。还注意,“模板”除了包含结构信息之外,还可能包含功能信息(图 22-7 中结合核苷酸的 β-α-β 超螺旋中的最后一个残基)。

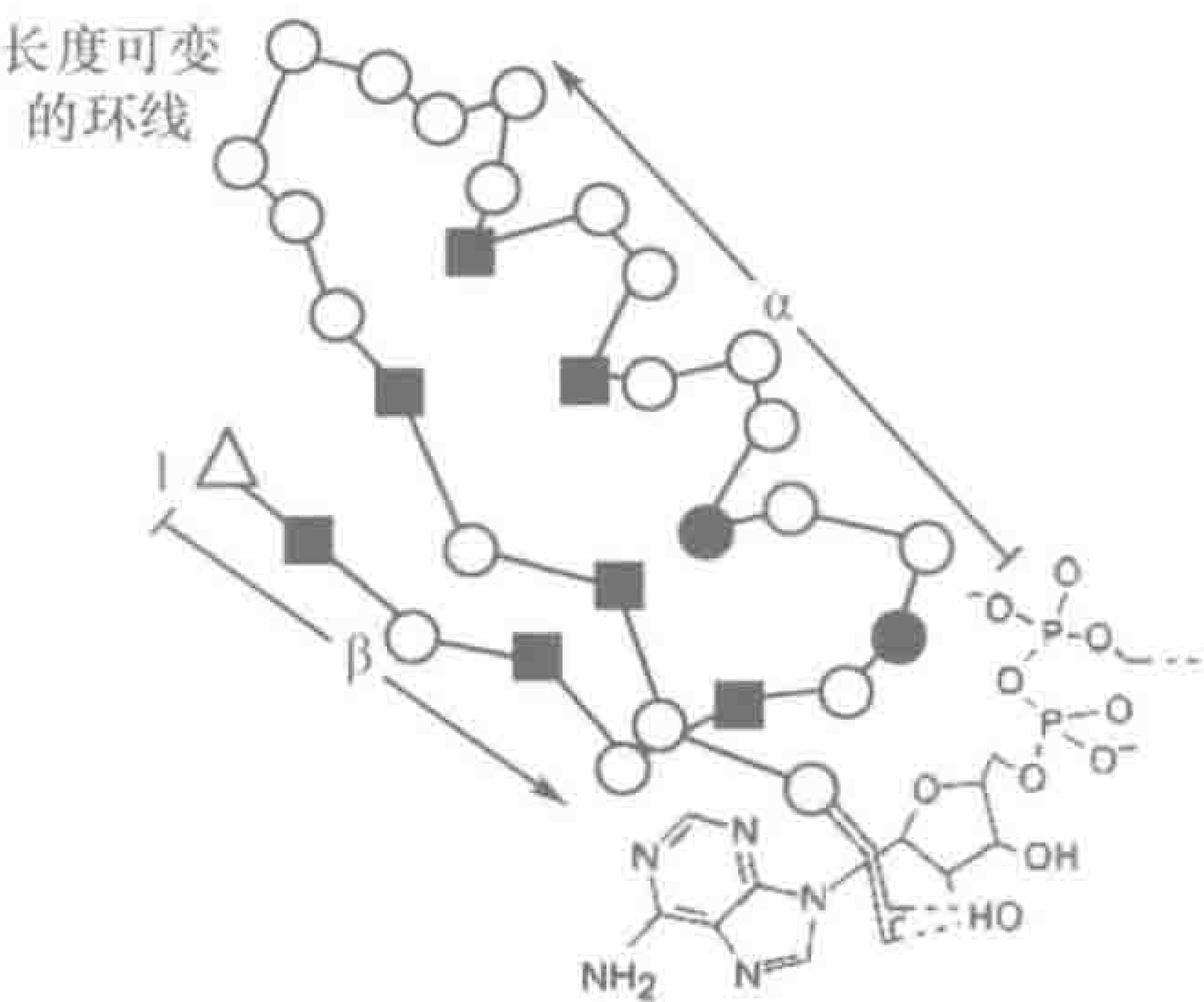


图 22-7 结合核苷酸的 β-α-β 超螺旋“模板”。正方形标示了通常由相对较小的疏水残基(Ala、Ile、Leu、Val、Met 和 Cys)所占据的位置;这是 β-α-β 超螺旋的疏水核。实心圆是只能由 Gly 占据的关键位置;该处有肽链的急剧转折。空心三角形标示的是模域(motif)β-α-β 的第一个位点,通常由带正电或含有偶极子的侧链所占据。模域 β-α-β 的最后一个位点(-)被能与配体(核苷酸)相结合的氨基酸 Asp 或 Glu 所占据。图片取自 Wierenga R. K., et al. 1986. J Mol Biol. 187:101-107,略有简化

■ 但是,让我们回到二级结构的预测上来。在开始时,让我们来讨论在“解折叠”肽链中的二级结构,即不考虑它们之间的相互作用。

如果已知各个分立的相互作用在 α 螺旋和 β 发夹的稳定性中的贡献,我们就可以计算这两种结构在链上任何位置的自由能。为此我们已经使用我们的程序 ALB 达 30 多年了。在其中一种工作条件下(“解折叠链”)它能够计算在链中每一个位置形成 α 螺旋和 β 结构的可能性,和它们在多肽和非球形(以及)蛋白质链中的含量,还可以在不同的温度、离子力和 pH 溶剂条件下进行计算。然后结果可以与实验数据相比较。例如,与同一多肽中根据圆二色性谱判定的二级结构的含量进行比较。图 22-8 显示,理论计算与实验吻合得不错。

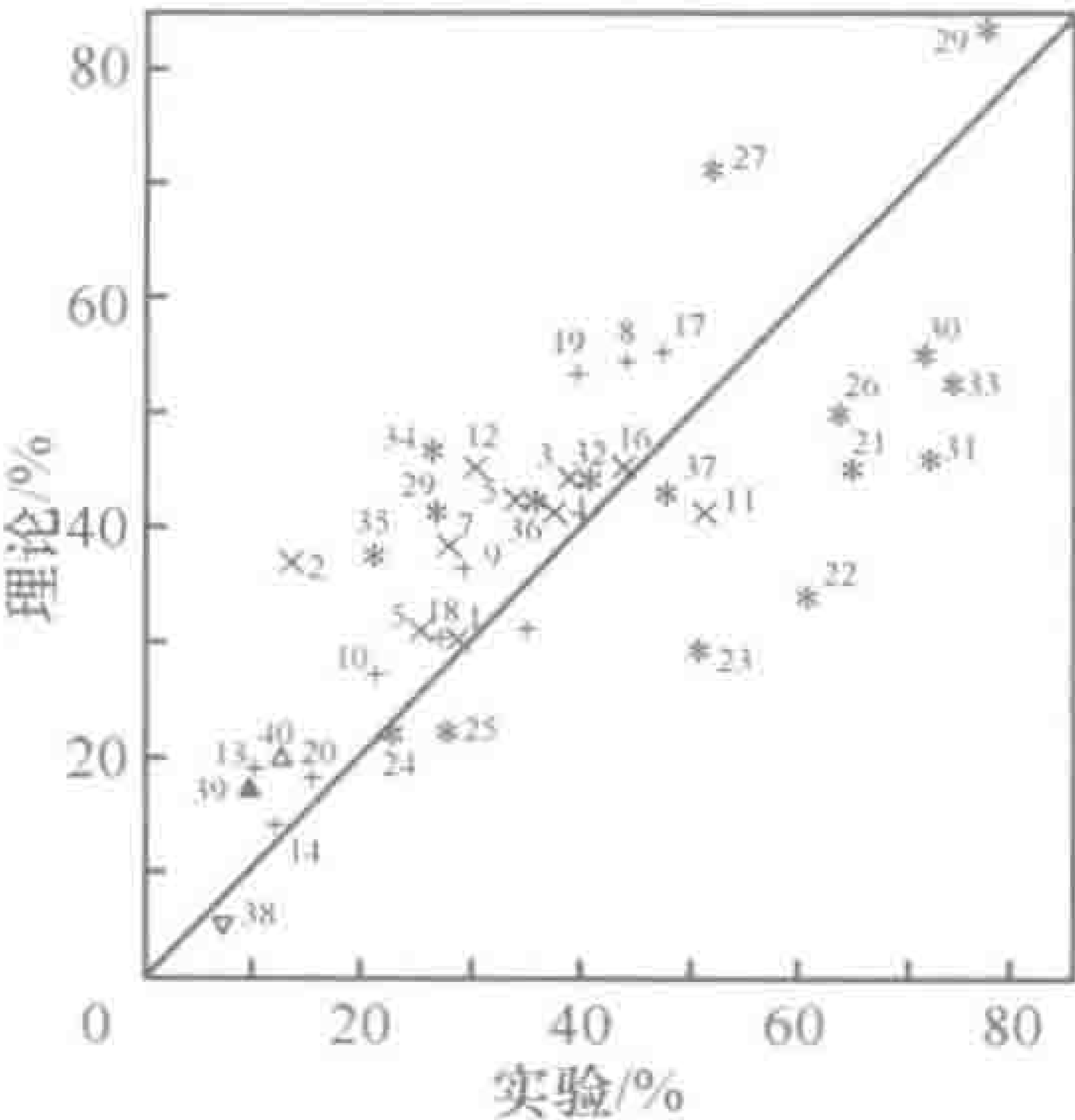


图 22-8 在温度为 0~5℃ 以及不同的离子力和溶剂 pH 下,理论上(由程序 ALB 对解折叠链进行计算)和实验上找到的几十个肽链的螺旋性。图片取自 Finkelstein A. V., Badretdinov A. Y, O. B. Ptitsyn. 1990. Proteins. 10: 287-299,略有修改

■ 在转到计算和预测球状蛋白的二级结构时,必须考虑到,这里对于已经存在于未折叠的肽链中的相互作用上,要加上每一个肽链区段与球体(**其构造我们还不知道**)的相互作用。更准确些,我们不知道它的细节构造,但是知道肽链区段以某种方式与蛋白质疏水核相衔接。在与核的相互作用的最简近似法中,可以通过二级结构与“疏水湖”(蛋白质链漂浮于其上)的相互作用来进行近似(图 22-9)。

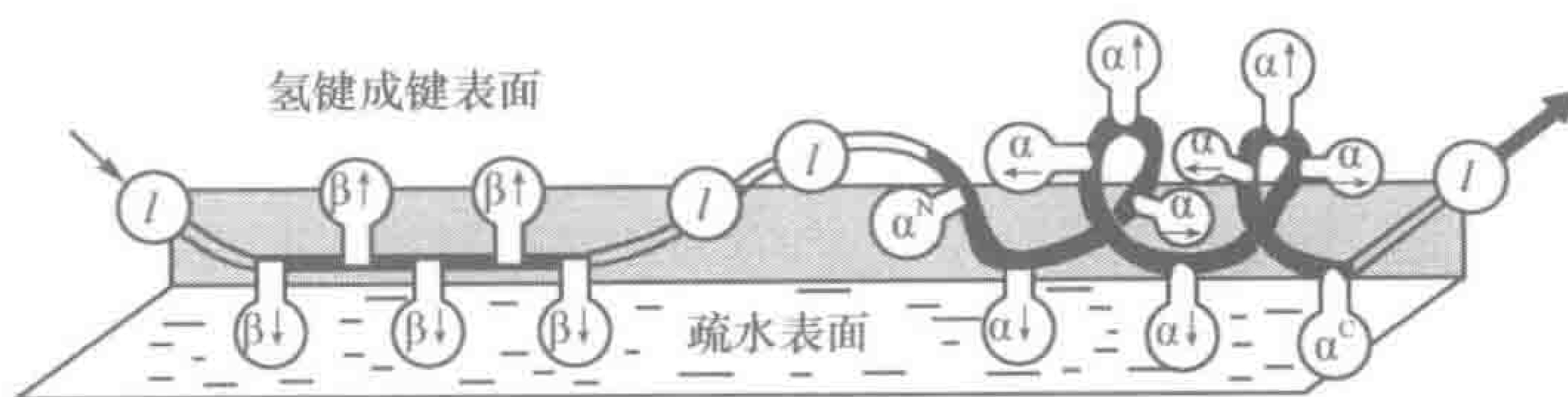


图 22-9　在用于模拟蛋白质球体的疏水表面上浮动的蛋白质链二级结构(β-β 串,*l*-环线,α-α 螺旋)(“浮木”模型)。该表面由疏水“湖”构成,其“湖岸”有助于与 β 区段结合氢键。模型考虑了侧链基团在不同的二级结构中指向湖表面(↓)、远离表面(↑)和沿着表面(→,←)的各种交替,以及在螺旋的 N 和 C 端上的作用(它们被综合并相应地归结到 N 和 C 端残基,α^N 和 α^C)。图片取自 Ptitsyn O. B. ,Finkelstein A. V. 1983. Biopolymers. 22:15-25,略有修改

一旦从实验中得知疏水相互作用的强度,而且从 α 螺旋和 β 结构的立体化学中知道朝向同一方向并因而能够同时与疏水表面相互作用的侧链基团在链中的交替模式,我们就可以计算出在蛋白质链的每一个位置形成 α 螺旋和 β 结构的可能性。ALB 也能做这些事[但是工作条件为“球态链”(globular chain)](顺便说一句,它的网址是 http://phys. protres. ru/resources. html)。

概率是在室温下进行计算的。为什么是在室温下?如果只挑选出 α 螺旋和 β 结构在肽链中的能量分布最好的,即在 0K 时计算所有概率,难道不更好吗?

当然,可以说,我们感兴趣的是蛋白质在(大约)室温下最稳定的结构,而维持它的,其中包括温度依赖性的疏水相互作用。

但是更重要的是,正是在这种温度(300K)下计算得的概率,略低于正常的蛋白质熔化温度(350K),所以这是最好的方法,用于挑出比较可能的、我们比较能确信的 α 和 β 区段,排除那些“可能性较低”、我们不能相信的。

要知道,本质上,我们是在努力只根据起作用的**部分**相互作用来预测蛋白质的二级结构:我们知道(大致上地)在该结构中的内部相互作用,但**不**知道(或只能非常粗略地估计)那些在所研究的肽链区段与球体其余部分之间的相互作用。而它们是很强烈的。

因此我们就处于这么一种状况,就是企图预测给定残基会处在蛋白质的内部或表面,但仅仅知道残基本身的疏水性,别的什么都不知道。而我们知道,这样的问题有解,但不是准确的,而只是**概率性的**。这个解包含在所观察到的残基在球体内部和表面之间的统计式中,我们已经知道,该统计式看上去是这样:

$$\frac{\text{处于内部的概率}}{\text{位于表面的概率}} \sim \exp\left(-\frac{\text{自由能}}{kT_C}\right) \qquad (22.1)$$

式中,T_C 为蛋白质统计学的特征温度,它(正如我们也记得的)稍微低于蛋白质的熔化温度。

回到蛋白质结构预测上来,在同样的 T_C 温度下,在仅仅知道二级结构本身的自由能的情况下,我们可以**概率性地**预测所讨论的二级结构的存在性。

而如果在这种情况下我们得到了可信的结论，而且完全或几乎消除了预测的浮动（祝你顺利！），则意味着，我们在计算中所采取的相互作用本身保证了足够稳定的蛋白质结构……

■ 图 22-10 展示了用 ALB-“球态链”程序对牛胰腺蛋白抑制子中的二级结构的预测结果。由于该计算考虑了二级结构内部的相互作用和它们与底物（模拟球体的其余部分）的疏水相互作用，而忽略了紧凑拼装，所以它其实是模拟了熔球的二级结构。但是正如我们知道的，天然球和熔球的二级结构相当接近。所得图片显示，计算得的有浮动的蛋白质二级结构的概率，甚至在结构性最强的位置，都远小于 100%。然而，与实验符合的是，α 螺旋在肽链的 C 端部分占优势，而 β 结构在中间部分占优势。因此螺旋态的概率峰处在较大的螺旋上，而 β 结构态的概率峰处在较大的 β 结构区段上。因此，根据这些概率峰（如果它们超过了根据经验设定的特定水平）也可以进行球状蛋白的二级结构预测。

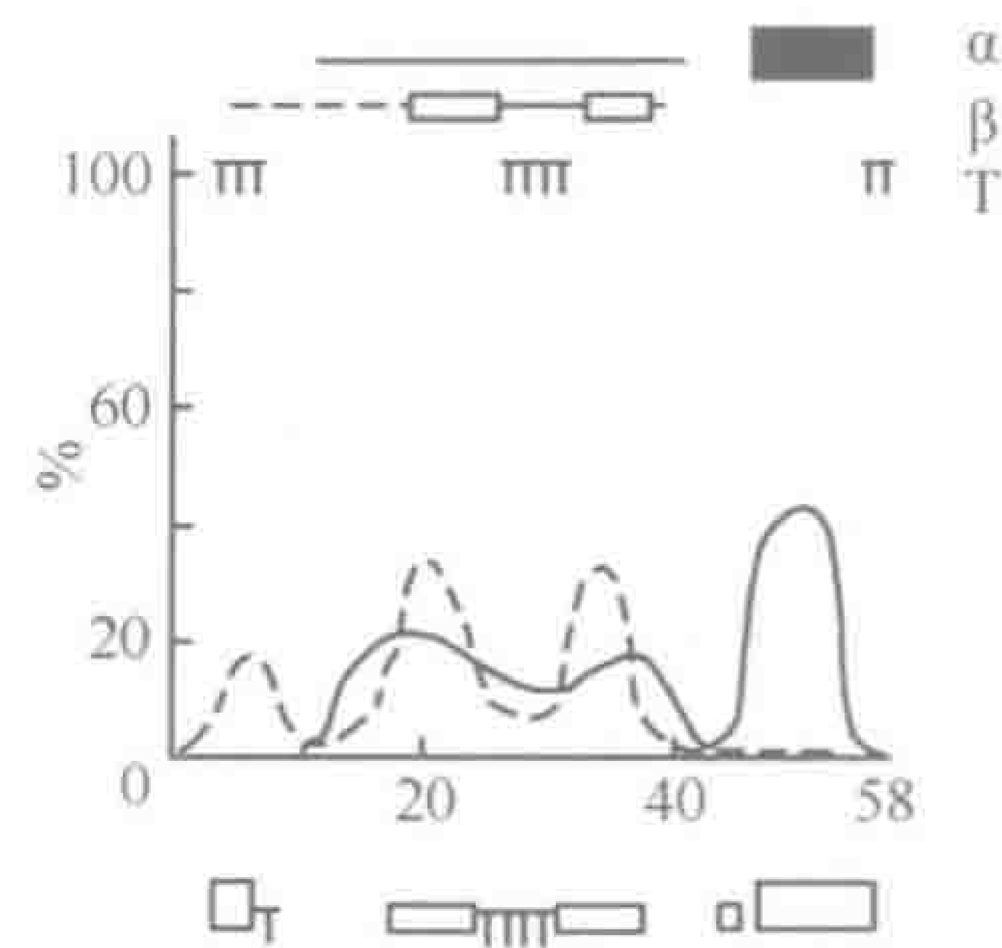


图 22-10　牛胰腺蛋白酶抑制子中的二级结构计算。横坐标：残基在蛋白质链上的编号，纵坐标：计算得的 α 螺旋态（—）和 β 结构态（---）概率。图上方显示了预测的最有可能的 α 螺旋（α）、β 串（β）和转折（T）。黑色长方形为可信的二级结构预测，空心长方形为不那么可信的预测，线条：给定的二级结构在原则上并不排除，但概率很小。图下方绘出了天然的 α 螺旋（高长方形）、β 串（矮长方形）和转折（T）在链上的分布

二级结构的预测总是具有概率性质的，这不仅是在 ALB 程序中，而且也在所有其他“预测”软件中，甚至当这些程序的目标（如通常所做的）是挑选出 α 和 β 区段在蛋白质链上的一个“最佳”分布时也是如此。

仅仅根据蛋白质的氨基酸序列，其 α 螺旋、β 结构和不规则环线的识别准确性约为 65%。

二级结构预测能不能改进？能。在 Rost 和 Sander 的“PHD”方法或 D. John 的“PsiPred”方法中（非常推荐它们；它们可以通过国际互联网访问），二级结构的预测不**仅仅**根据单独的氨基酸序列，而且（只要可能就总是如此）根据同源序列的集合进行。在这种方法的结果中，随机误差被消除，而且预测变得更准确得多（平均上能达到 75%，而不是 63%~65%）。

因此，尽管精度有限，二级结构预测实际上已经变成了蛋白质一级结构研究中的日常方法。

根据蛋白质的氨基酸序列来预测空间结构的问题要复杂得多，情况更糟。对此我们将在下一讲中讨论。

第23讲

根据氨基酸序列预测蛋白质空间结构的方法概述。蛋白质结构数据库。远同源链的通用折叠预测降低了蛋白质结构识别中的不确定性。结构基因组学和蛋白质组学。生物信息学。蛋白质工程和蛋白质设计。蛋白质结构设计中的初步成功。

让我们现在来到一个发展迅速的新领域:蛋白质链的空间折叠预测。

■ 空间结构的计算常常基于先前的蛋白质链二级结构预测。虽然这条方法并不是严格地理由充分:因为三级结构可能影响二级结构,但有时它能取得成功。

图23-1展示了球体中的二级结构计算。它是1985年在干扰素上进行的。看得出,在其中α螺旋占优势,尤其是在肽链的N端部分,该处(正如已从实验得知的)存在有干扰素的功能域。在本例中,这些螺旋被如此确信地(而这并非总能如此)预测到,而且在不同的并无高度同源的肽链中如此一致,以致我们在1985年也尝试用这些螺旋拼接成域。

获得的复合体是由三个大螺旋和一个小螺旋构成的螺旋束,显示在下图中[图23-2(a)]。在这个预测完成的五年之后,最终获得了β干扰素的X射线结构[图23-2(b)],而且它的N端域的X射线结构与我们预测的结果相当精确地相吻合。

干扰素的N端域的结构是最早比较成功地在实验之前被预测到的蛋白质分子结构之一。

需要理性地强调有助于成功预测干扰素空间构造的因素:α螺旋本身在这里被非常确信地预测到了;而且它们在远同源肽链中的分布相同,这有助于确信所获得的预测结果。然而如此确信并且在同源物中重复的二级结构预测在实际上并不能经常碰到。更典型的是有一些差异,就像在干扰素的C端域中一样(图23-1),正是由于这些在二级结构预测中非常典型的微小差异,对干扰素的C端域并不能达到同样的肽链折叠预测。

对干扰素构造预测的成功还有一个非常重要的原因。在我们寻找预测得的α螺旋的最佳拼装时,我们可以根据对α螺旋复合体的先验分类(对此我已经讲过)(在那时它已经由Murzin和A. V. Finkelstein研制出来了)列出所有可能的拼装,并进行合理而详尽的轮换搜索。

■ 蛋白质结构的预测问题可以作为选择问题来提出:从大量可能的空间结构中挑选出对给定的氨基酸序列最稳定的那一个。那么从哪里能得到"可能的空间结构"呢?

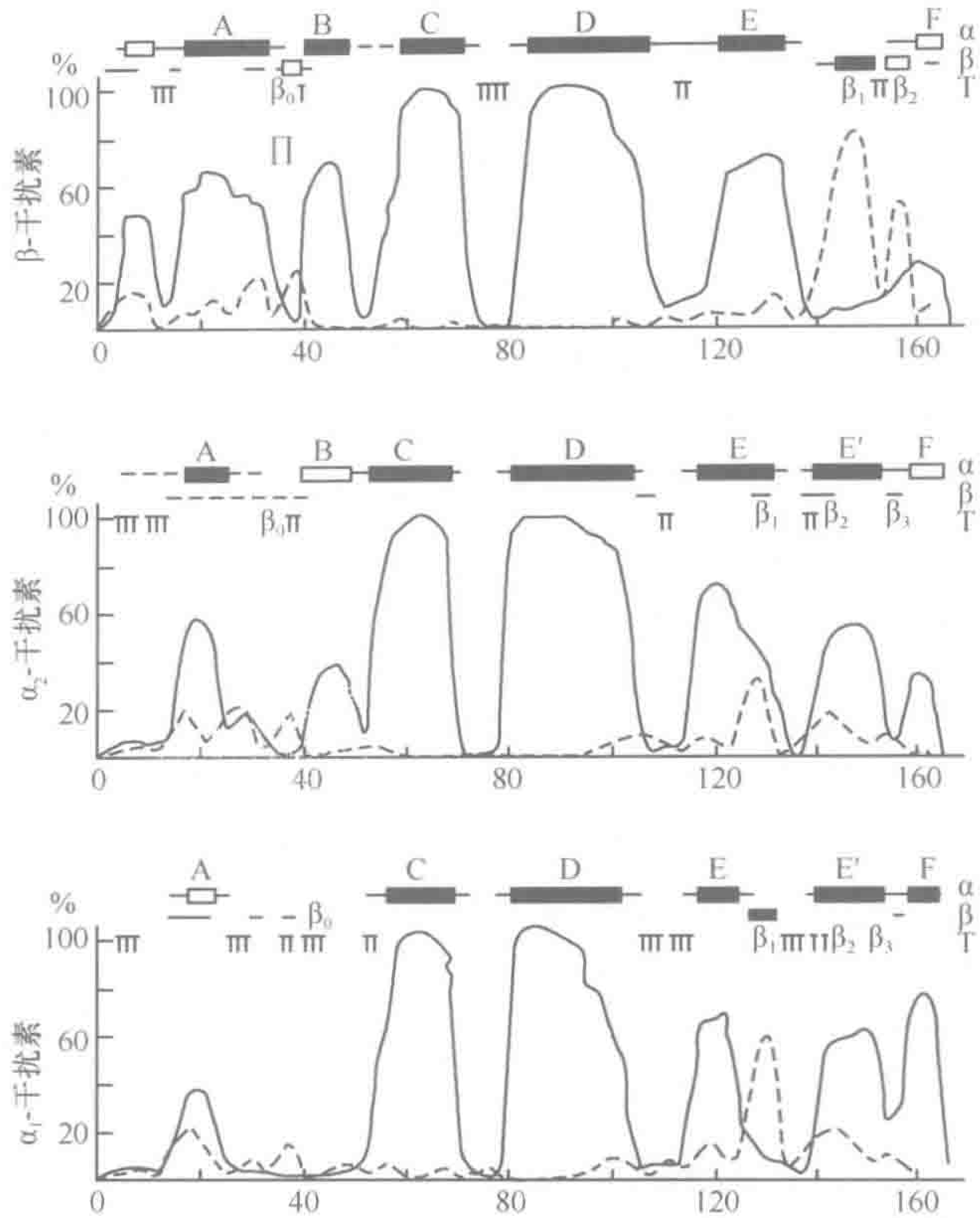


图 23-1 几个干扰素的二级结构计算。横坐标为蛋白质链上的残基编号,纵坐标为计算得的它们的 α 螺旋态(—)和 β 结构态(---)的概率。上方显示了概率最大的 α 螺旋(α)、β 串(β)和转折(T)的预测。黑色长方形为确信的二级结构预测,空心的为不那么确信的预测,直线:给定的二级结构在原则上并不排除,但概率很小。形成 N 端域的肽链区段在图下方以着重线标出。图片取自 Ptitsyn O. B., Finkelstein A. V., A. G. Murzin. 1985. FEBS Letters. 186:143-148

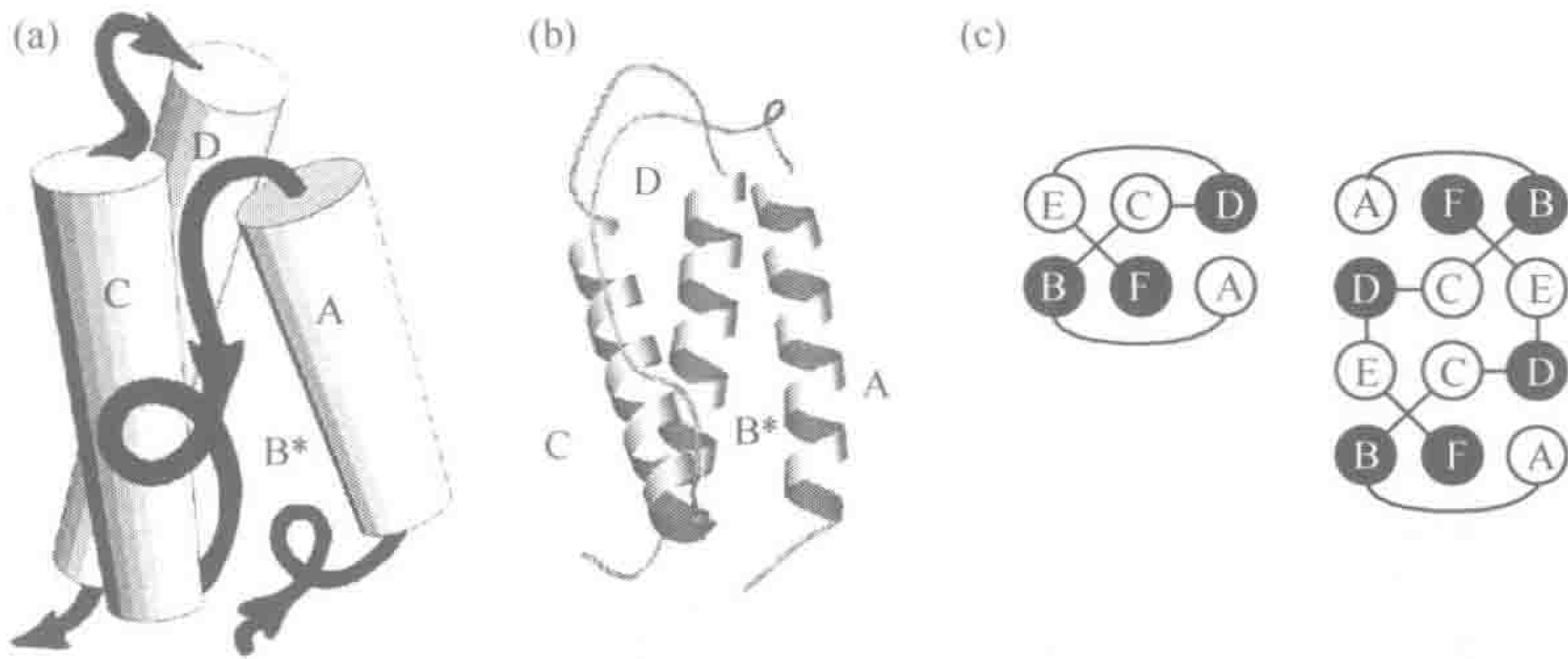


图 23-2 (a)干扰素的 N 端结构域的链折叠预测,1985 年完成(Ptitsyn, Finkelstein, Murzin. 1985. FEBS Letters. 186:143)。三个大 α 螺旋(A,C,D)显示为圆筒;除此之外,也预测到可能存在单独的螺旋圈 B。(b) β 干扰素的 N 端域的 X 射线结构(C 端域未绘出),于 1990 年解析得到(Senda, et al. 1990. Proc Jpn Acad Sci, ser B Biopys Biol Sci. 66:77)。图(b)中的方向与预测模型(a)相同,且其中也给出了与图(a)中相同的大螺旋名称(A,C,D)。在 β 干扰素中,B * 区域是螺旋形的,但不是 α 螺旋,不过在与它同种的 γ 干扰素中有较小的 α 螺旋(Ealick, et al. 1991. Science. 252:698)。(c) β 干扰素(左)和 Ealick 等人得到的 γ 干扰素(右)的拓扑结构。γ 干扰素由两个亚基构成。注意在 γ 干扰素的两个亚单位之间“交换”了 C 端小块(螺旋 E 和 F)

最简单的,实用主义的回答是,从蛋白质结构数据库(Protein Data Bank, PDB)①中。于是,在本质上,问题不是关于蛋白质空间结构的"预测",而是"识别"。但是在这里有它的优点:一旦"识别"到蛋白质,我们可以期望同时识别到它的功能,而要知道,这正是生物学家通常也感兴趣的。

当然,在蛋白质结构数据库中尚未包含所有的可能结构,但是可以期望,那里大致包含了超过一半的存在于自然界中的蛋白质链折叠成域的所有折叠模式。这一期望(由 Cyrus Chothia 所证明)的基础在于,最新被解析的蛋白质(更准确些,是它们的域)越来越显得与已解析的蛋白质相似。按照 Chothia 的估算,根据已解析结构的数量和肽链折叠模式数量进行外推,自然界中存在有大约 1500 种折叠模式,大部分(而且最"流行"的)已经知道。

■　因此,在预测与已解析的蛋白质没有明显同源序列的蛋白质的结构时,可以尝试从数据库中逐个地取出空间结构,把该蛋白质链比对在每一个结构上面(可以有一些外侧环线,如图 23-3 所示,或者甚至可以改变链穿过蛋白质结构骨架的行进通路的次序),并看看其中的哪些空间结构对该蛋白质链能给出最优的能量。在这种情况下我们应该准许肽链或沿结构骨架行进,或发出外侧环线,或"缩短"在骨架中的外侧环线,如果这能优化能量的话。

这种方法被称为"穿线法"(threading method)。它是由 A. V. Finkelstein 和 B. A. Reva 在 1990 年提出来的,David Eisengerg 和他的研究组在 1991 年也独立地以更简单、更方便、尽管在物理学上没那么正确的变形形式提出该方法。现在,穿线法已经成为根据与"旧"蛋白类比来识别"新"蛋白的结构的最流行方法。

基本上,穿线法与同源序列搜索相似,只是在这里不是两个一级结构("新"和"老")进行"比对",而是"新"一级结构与"老"蛋白质的骨架空间结构进行比对。

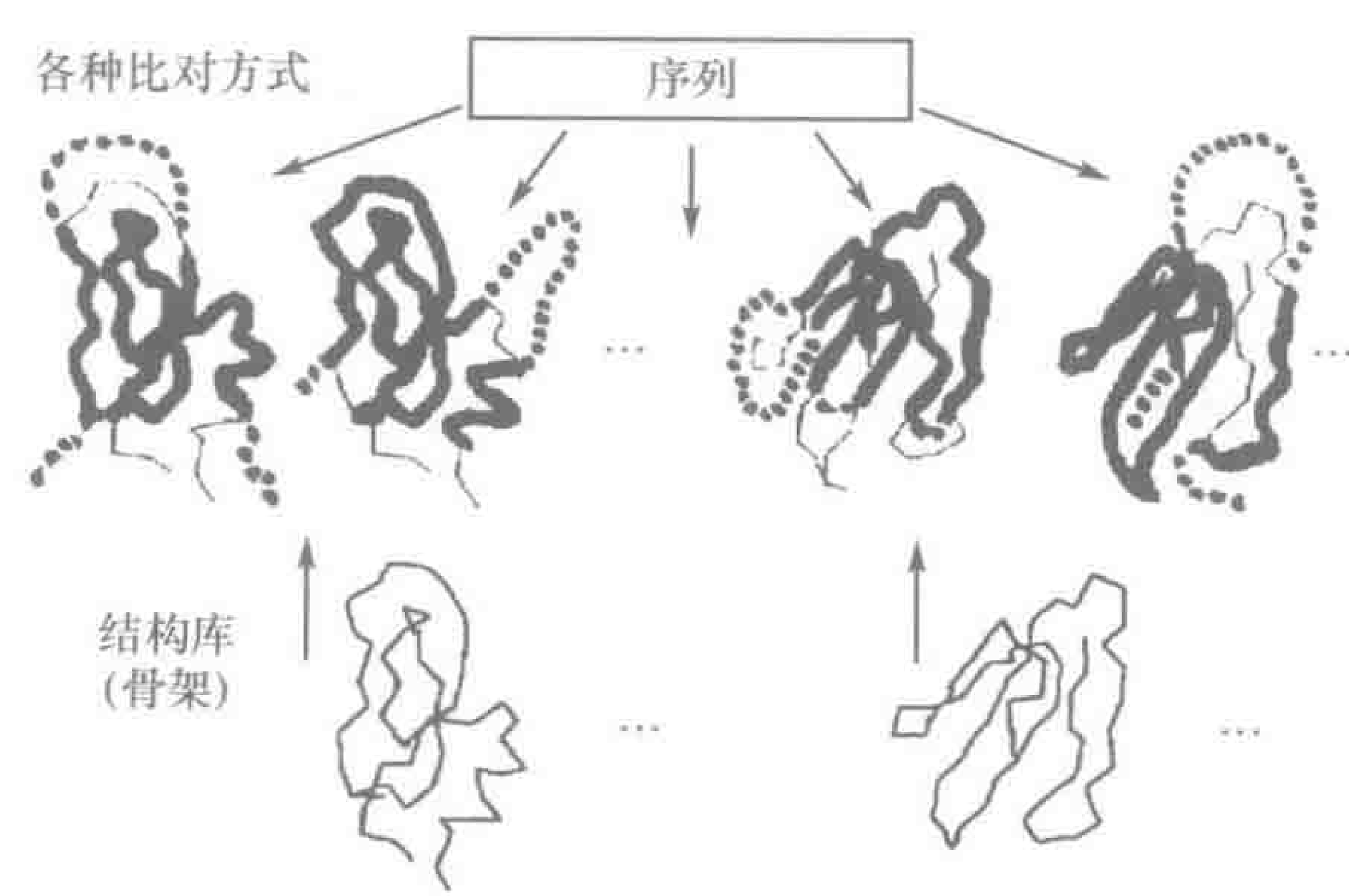


图 23-3　展示根据蛋白质结构数据库来研究序列的"穿线法"思想的示意图。粗线条表示那些序列沿骨架行进的区段,虚线表示伸展出的外侧环线

---

① 蛋白质结构数据库(PDB)是一个国际性的分子结构数据库,收录了所有三维空间结构已被实验测定的蛋白质分子(和其他分子)的结构信息。网址:http://www.rcsb.org.

■　在讲述结果之前,我想强调“穿线法”的两个原则性问题。本质上,在任何“预测”方法中都会出现相似的这样或那样形式的问题。

第一,我们研究肽链区块(甚至那些位于骨架上的)的构象时有很大的误差:因为我们不考虑侧链基团的构象(在“新”和“老”肽链上不同),而实质上正是它们(侧链基团)发生相互作用。更进一步,我们不考虑所有外侧环线的构象。估算显示,在穿线法中我们大约只考虑了蛋白质链中的相互作用的一半,而另一半根本完全不考虑。这意味着,我们再一次被迫根据蛋白质链上的部分相互作用对蛋白质结构进行判断。这意味着,我们的预测再一次可能只具有概率性质。

第二,如何逐个检视所有比对并从中找出更好的,或者最好的?因为它们(可能的比对)实在是太多了……长话短说吧:为此已经发展了强大的数学计算方法,但讲述它们会占用太多的时间。也许,提到这些方法的名称就够了,这也许对你们中的某些人有用(当然,对大部分人没用)。就是这些:动态规划及其变体:一维系统(分子链)的统计物理,用于肽链穿过骨架的穿线法计算;自洽场理论:用于计算在骨架上每一点上作用在肽链上的分子场;蒙特卡罗方法随机能量最小化;以及分支和边界方法的各种变体,如此等等。还有:对于比对的直觉估计法!在计算机时代看上去这已经过时了,但正是这种方法让 A. G. Murzin 预测了许多“新”蛋白质的结构,在这些独特的蛋白质棋局中经常胜过最好的计算机程序。

讲课人:上面这句话是在本书的 2001 年第一版中写下的。好玩的是,在过去的这些年里,尽管计算机有巨大的进展,“蛋白质棋局”的情况却没有本质的变化(见文献 Predicting protein structures with a multiplayer online game. Nature (2010) 466:756-770 和计算机游戏 Foldit: http://fold.it/portal/node/)。

■　不管怎么样,获得的结果怎么样呢?作为例子我展示一下在复制终止蛋白质中用穿线法得到的肽链折叠预测结果。通过把这个“新”蛋白质的肽链对所有已知的三维蛋白质结构进行穿线,Sippl 及其在澳大利亚萨尔茨堡的研究组证明了组蛋白 H5 与该链的“亲源性”最大(图 23-4)。

通过这种方法,他们正确地(尽管氨基酸序列的相似性很低)识别出复制终止蛋白质链的折叠模式。确实,识别(通过理论上的穿线法)得的该蛋白质链在组蛋白 H5 的结构上的比对与这两个分子的三维结构的直接比对有相当大的偏差(图 23-5)。这再一次表明,我讲到过的误差:稳定性参数中的误差,准确的环线构象未知,等等,让人无法挑选出**单个**唯一的最佳结构,而在实践中只是让人找到比较狭窄的较好结构的集合。这样的“较好”空间结构的集合可以相当期望地挑选出来;而从它们之中挑选出看上去(根据计算)“最好的”,则是偶然的。天然结构处于“较好”结构集合中的**某处**,它大约对应着预测到的(“最佳的”)结构,但这就是实际上全部可说的了。

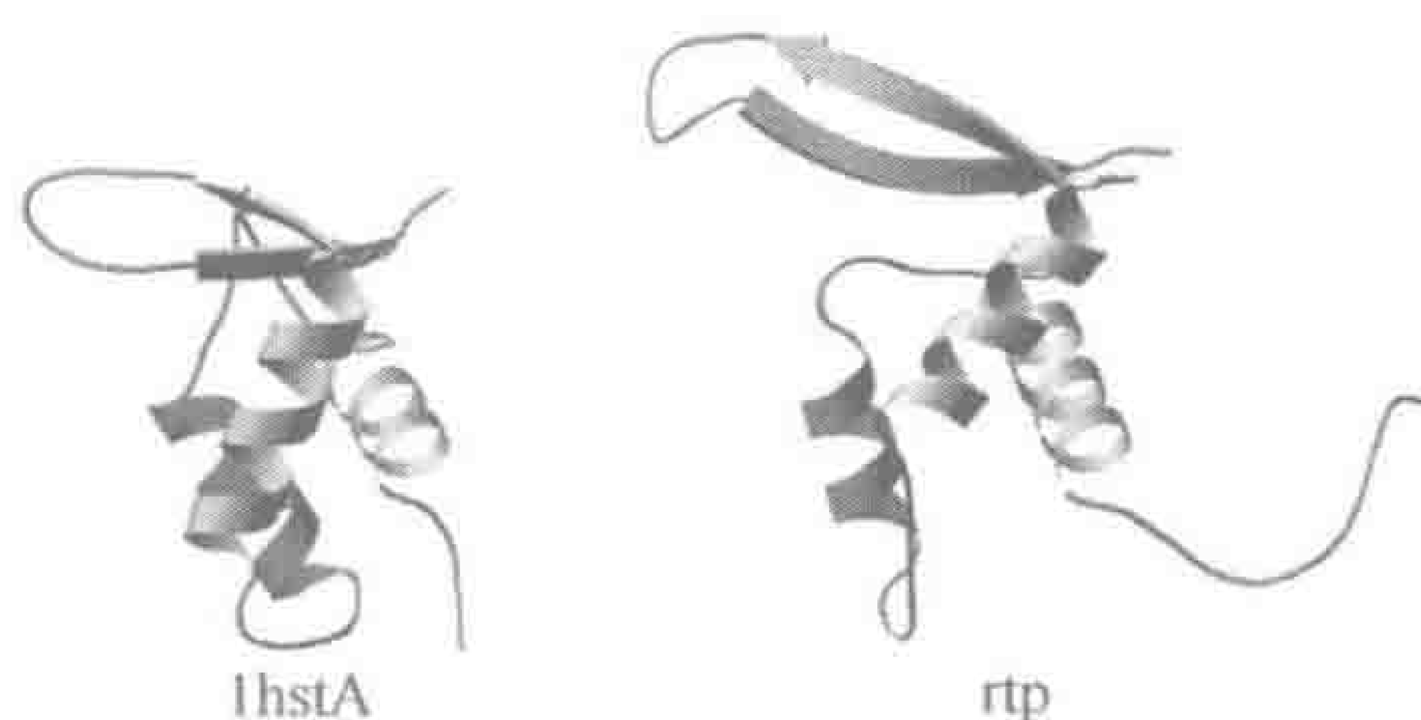


图 23-4 鸡组蛋白 H5(1hstA,左)和复制终止蛋白(rtp,右);在后者中没有画出 C 端螺旋,它在组蛋白 H5 中没有相似物。用“穿线法”在这两个蛋白质中的肽链折叠的相似性预测是在预测方法的“双盲”竞赛(CASP-1994)①过程中完成的(Flöckner H., Braxenthaler M., Lackner P., et al. 1995. Proteins. 23:376-386)。在所画出的结构中的 65 个等价 C^{α} 原子的均方偏差为 2.4Å

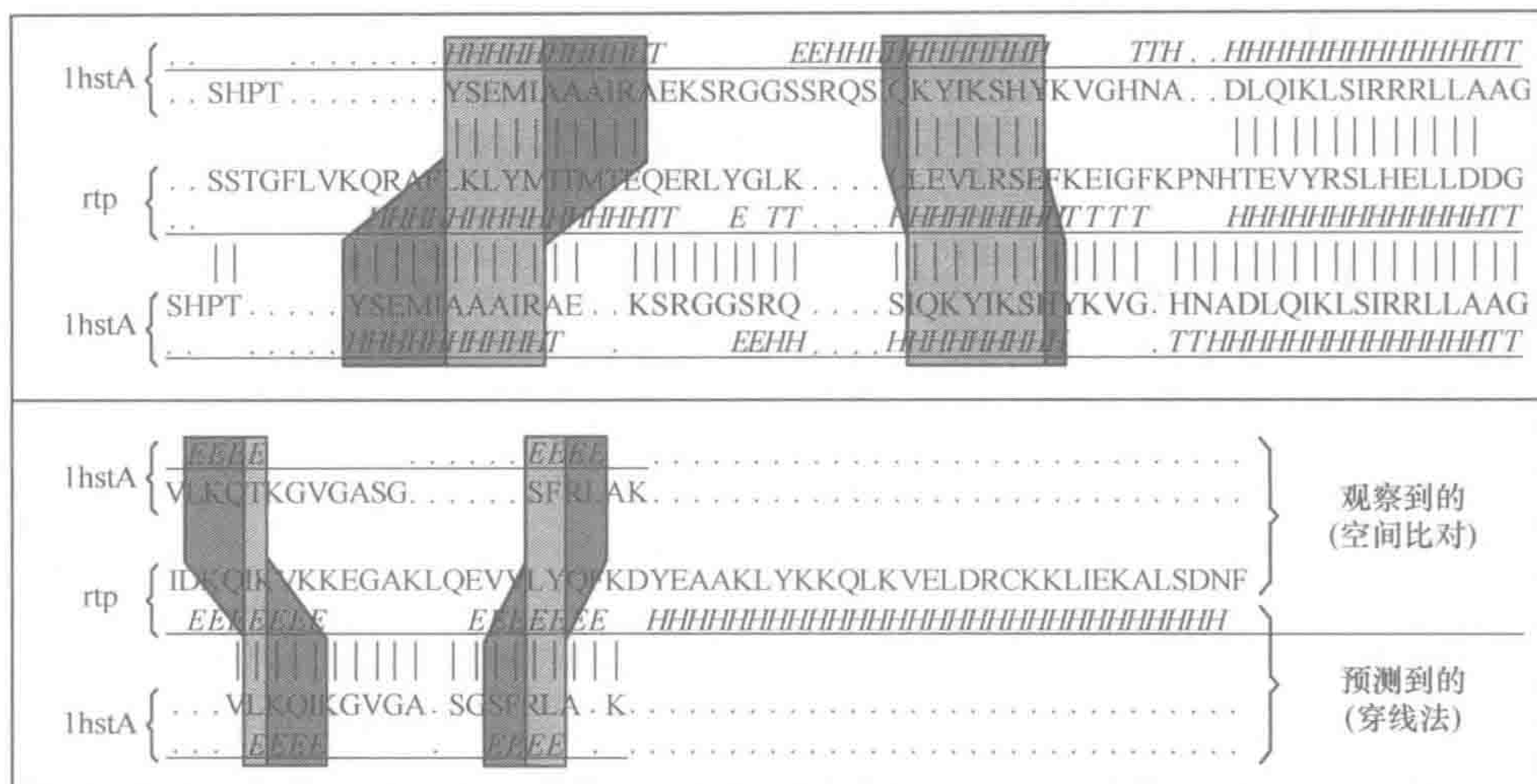


图 23-5 组蛋白 H5(1hstA)序列与复制终止蛋白(rtp)序列根据 X 射线结构数据进行的比对(“观察到的”,上面一对序列),以及根据穿线法预测进行的比对(“预测到的”,下面一对序列)。青色区域显示了这两个比对的偏差平移。它们很容易根据列出在比对的最顶行和最底行的组蛋白 H5 的二级结构的不同平移看得出来。二级结构写在序列下方(标志:*H* 为 α 螺旋,*E* 为 β 串,*T* 为转折)。在一级结构比对和二级结构比对中引入的空位以点表示。图片取自文献:1995. Protein structure prediction by threading methods: evaluation of current techniques. Proteins. 23:337-355(该文中总结了 CASP-1994 的结果),形状上有微小修改

当然,我举出的只是“好”例子:预测成功的例子。不成功的例子多得多,而这可以理解:因为在预测时需要从无数的变体中挑选出一种或几种。因此,如果有一半情况(现在最成功的研究组的工作有这个水平)能击中这样无限小的目标,就是巨大的成功了。

---

① CASP(Critical Assessment of Techniques for Protein Structure Prediction,参见网站:http://predictioncenter.org/)是世界范围内对蛋白质结构预测的技术方法的比赛,号称是此领域中的奥林匹克比赛。第一届是在 1994 年开始的,以后每两年举行一次,一般第一年预测,第二年给出全面的评估。

■ “穿线法”现在已经成为蛋白质序列三维结构识别的工作工具。极其重要的是,已经形成了配方:只要这样、这样和这样做,结果你就能得到一个或几个结构,有相当高的可能性,在它们当中会有一个结构与所研究序列的三维折叠相似。

■ 在这一背景以及在“根据同源物”进行蛋白质结构识别的进展(尤其是诱人的可以检索、运行多种程序并总结其运行结果的生物信息海量服务器)背景之上,通过仿真蛋白质折叠而进行结构预测的成功事例总的来说,目前看来是比较少的。不过这里还是给出几个相当突出和有趣的工作。

我想特别提到 David Baker 的使用折叠仿真寻找肽链稳定折叠的方法。这一方法体现为程序包 Rosetta,它不仅成功地应用于蛋白质结构预测,也成功地应用于将在本讲末尾讲到的蛋白质设计。在最开始时,对结构待预测的蛋白质链的每一个短片段生成了同源物的集合(从已知的空间结构中)。然后由这些同源物片段的空间结构组装出所研究蛋白质的大量可能的空间结构,再然后根据能量计算而剔除不合格的结构,而挑出较好的结构。通过这样做,Baker 在预测方法的“双盲”测试(CASP)中不仅能“识别”在 PDB 数据库中的蛋白质链折叠,而且由从头算起(*ab initio*)预言了几个“新的”、尚未被观察到的蛋白质链折叠。

还必须提及不久前 V. Pande 研究组 (2010. J Am Chem Soc. 132:1526-1528)尤其是 D. Shaw(2010. Science. 330:341-346)的工作。他们能够追踪一些蛋白质解折叠链的折叠,尽管目前只是微小的蛋白质在几十微秒的时间内的折叠,结果不仅证明了它们能达到自身的天然结构,而且能正确地估算它们的折叠速度。

■ 能否消除在蛋白质中的相互作用近似估算中的误差效应(参见习题 23.1),从而改进蛋白质结构的识别和预测? 可以,如果利用上同源序列的话,即预测蛋白质结构不仅根据**它的**氨基酸序列,而且根据同源序列**集合**。使用同源序列能改进蛋白质二级结构预测。该方法广泛并成功地应用于 RNA 二级结构预测。在蛋白质链的空间折叠识别中也开始应用,在这种方法中认为那些对所有同源序列都稍微好一点的折叠是更好的。而且看来相似性为 ~40% 的同源序列是最优的:它们还几乎保留着完全相同的空间结构,但是已经按不同方式编码,这就使得在所有预测误差“噪声”中能识别出“信号”(结构)。不过,在这种方法下我们得到的不是给定肽链的准确结构,而是一些通用的,即对所有同源序列都适用的近似结构。

■ 我今天的讲述内容涉及的只是球状蛋白的三维结构预测。而对纤维蛋白和膜蛋白能讲些什么呢?

在纤维蛋白的序列中(不用任何计算机)能看出显示其二级结构特征的周期性,而且它们是如此明显,以致识别这些蛋白质的二级结构通常都不困难。而且在这个基础上成功地预言了它们拼装成纤维的几种可能形式。

膜蛋白结构预测发展得不如球状蛋白:我们还只知道非常少量的膜蛋白结构。

识别这类蛋白的膜内部分是相当容易的：它们是连续（更准确些：几乎连续）的疏水模块，长度由膜厚度所规定。然而在蛋白质中的这些模块形成折叠的原理还没有完全弄清楚（虽然主要的已经清楚；我们记得：膜蛋白的“内部”结构不像球状蛋白那样是疏水的，而是极性的；那些疏水基团是与脂质膜相接触）。总之，膜蛋白空间结构的预测工作还刚刚开始。

■ 总结一下。

从物理学的观点来看，蛋白质链的三维折叠预测就是寻找所研究的肽链的最稳定结构。这种预测在原则上是可能的，而且有时候能达到完全满意的结果。然而它总是复杂的，这与蛋白质链的可能折叠的清单不够详尽和蛋白质链中大量相互作用的信息不够精确有关，也与能量参数中的误差等等有关。

结果就是，对与已知结构没有同源序列的肽链，主要是根据其序列而进行折叠预测时，单一可靠的预测至今不大可能，不过可以挑出几个最可能的肽链折叠模式，并（这很重要）可以剔除数量巨大的其他折叠。而且如果考虑到额外的信息：尤其是，同源蛋白质的肽链序列以及它们的可能折叠，它们的活性中心，等等。这样的预测可以准确化。

■ 在实用主义的角度上，蛋白质三维结构的预测问题现在与其说是预测，不如说已经变成了**识别**。它变成了这样：所研究的蛋白质序列的空间结构是否与某个已知的蛋白质空间结构相似？如果**是的**，则所研究的序列在该结构中**究竟**是如何排布的？

识别的策略是受限的（因为并非所有可能的蛋白质空间结构都已知，而只能识别已知的），相对于先验性的（且不被已知结构集合所限制）**预测**，它具有一个毫无疑义的优势。一旦识别了蛋白质结构，我们常常可以猜测到它是干什么的；而结构预测，它本身没有与任何已知结构的相似性识别，在这方面不能给出任何暗示。

■ 应该指出的是，用于预测（更准确些：识别）蛋白质结构的方法只有一部分是基于物理学的。大部分方法都试图识别给定序列（且经常地，它在基因组中的环境）与已知蛋白质的序列（和基因组环境）的相似性。

所有的“识别”方法都基于生物大分子聚合物的结构数据库和序列数据库。这些数据库包含了海量的信息，越来越复杂，并壮大了一种趋势：人们可以从这些数据库获取所有必需的信息：可能结构的集合、它们中各种不同相互作用的能量、同源物家族、可能功能的集合和对这些功能适用的活性中心。这被称为“生物信息学”；这个科学分支发展得非常迅速，已经投入了大量的资金。数据库的完善也花费了不少资金。你们都知道“人类基因组”计划，它要求建立人类和一系列其他生物体的所有基因（总共有几十万个基因）的一级结构。在人类染色体测序部分，即确定核苷酸序列，这个工程实际上已经完成。好几个更复杂的事情与对这些染色体上的基因进行鉴定和注释有关。鉴定真核生物 DNA 上的基因是一项不平凡和

困难的任务:因为(请回忆剪接,尤其是可变剪接)需要找到组装成编码蛋白质的基因的所有小块。然而,人们希望这个任务将在几年以后得到解决。

如下几个大规模的生物信息学计划已经开展起来了。

(1)“蛋白质组学”计划要求识别所有编码蛋白质的基因,且进一步地认识每一个编码蛋白质的基因在基因组上下文中是做什么的:识别该蛋白质的生物化学功能(基因是编码氧化还原酶?或核酸酶?或?),该蛋白质在细胞中的作用(参与核苷酸代谢?或调节?或?),以及该蛋白质与其他基因的产物之间的联系。

(2)蛋白质链折叠的识别可能作为一个工具而服务于这项计划。因为,在最终结果中,生物学家感兴趣的正是蛋白质的功能,而结构和序列服务于这一需求。

(3)另一个是“结构基因组学”计划,要求对每一个蛋白质家族至少解析(用X射线或NMR)一个样本的三维结构,这一计划应该在大约10年时间内总共给出数10万个不同的蛋白质结构(超过100亿美元的经费)。当这一计划完成时,以从头算起(*ab initio*)方法对蛋白质结构的预测将变得多余:我们感兴趣的序列的空间结构将只要通过同源物进行识别就行了。

(4)最后是宏大的“个体基因组”计划:从单个人体的基因组到伤害人体的肿瘤、细菌和病毒的基因组……

■ 在本讲的结尾部分我想简要地讲述一下**蛋白质工程**,即对已存在的蛋白质进行定向修改,或**蛋白质设计**,即新蛋白质的结构设计。

寡核苷酸合成和DNA重组技术使人们能够获得自然界不存在的蛋白质的基因;X射线和NMR技术使人们可以看到蛋白质的三维结构;而高性能的计算机和计算机图形学使人们可以与这些空间结构进行交互对话,在它们里面进行改变并估算其影响。这些方法结合起来,它们就相应地成为分子生物学新领域(蛋白质工程)的“手”、“眼”和“脑”。它的战略目标是建立起能获得具有事先给定的功能和结构的蛋白质的知识和方法。这些研究对新药物和催化剂的结构设计、对纳米技术等的前景无论如何高度评价都不过分。

定向蛋白质工程实验已经对一系列基础性问题给出了答案。已经证明,在选出空间结构的过程中远非所有细节都符合紧凑拼装,蛋白质结构经受不起大量点突变的检验;而且环线未必符合这种挑选:如果从蛋白质链上替换或者剪除编码环线的区段,则球体上的“伤口”通常能被拉合。我们已经讲过,定向引入点突变能给出大量有关蛋白质球体和在其自动折叠过程中的过渡结构的能量信息;而它们如何应用于研究蛋白质活性,我们将在下一讲中讲述。

还应该提及的是蛋白质工程在药理学和化学中的工业应用:通过引入定向突变,可以获得稳定性得到增强(或降低,这取决于需求)和催化活性经过改造的蛋白质。

■ 用于改造蛋白质的蛋白质工程方法的要点显示于图23-6中。假设我们想要在蛋白质结构中引入某个位点,或仅仅只想使该结构稳定化(作为一个简单例子,

图 23-6 中显示的是引入经常出现在蛋白质活性中心内的位点 Asp-His)。则首先在"老"蛋白质的数据库中(它们的结构已经用 X 射线或核磁共振测定出来)找出合适的蛋白质以及其中的氨基酸残基(图中的 Z 和 Y),能满足在作替换 Z→Asp 和 Y→His 之后,能够给出具有 Asp 和 His 的所求构象的所求位点。然后在所选出的"老"蛋白质的基因中,我们可以把编码 Z 和 Y 的密码子相应地替换为编码 Asp 和 His 的密码子,就得到了含有所求位点的"新"蛋白的基因[当然,然后需要检查该位点能否工作,并弄清(借助于 X 射线或核磁共振)它是否获得了所需的构象]。

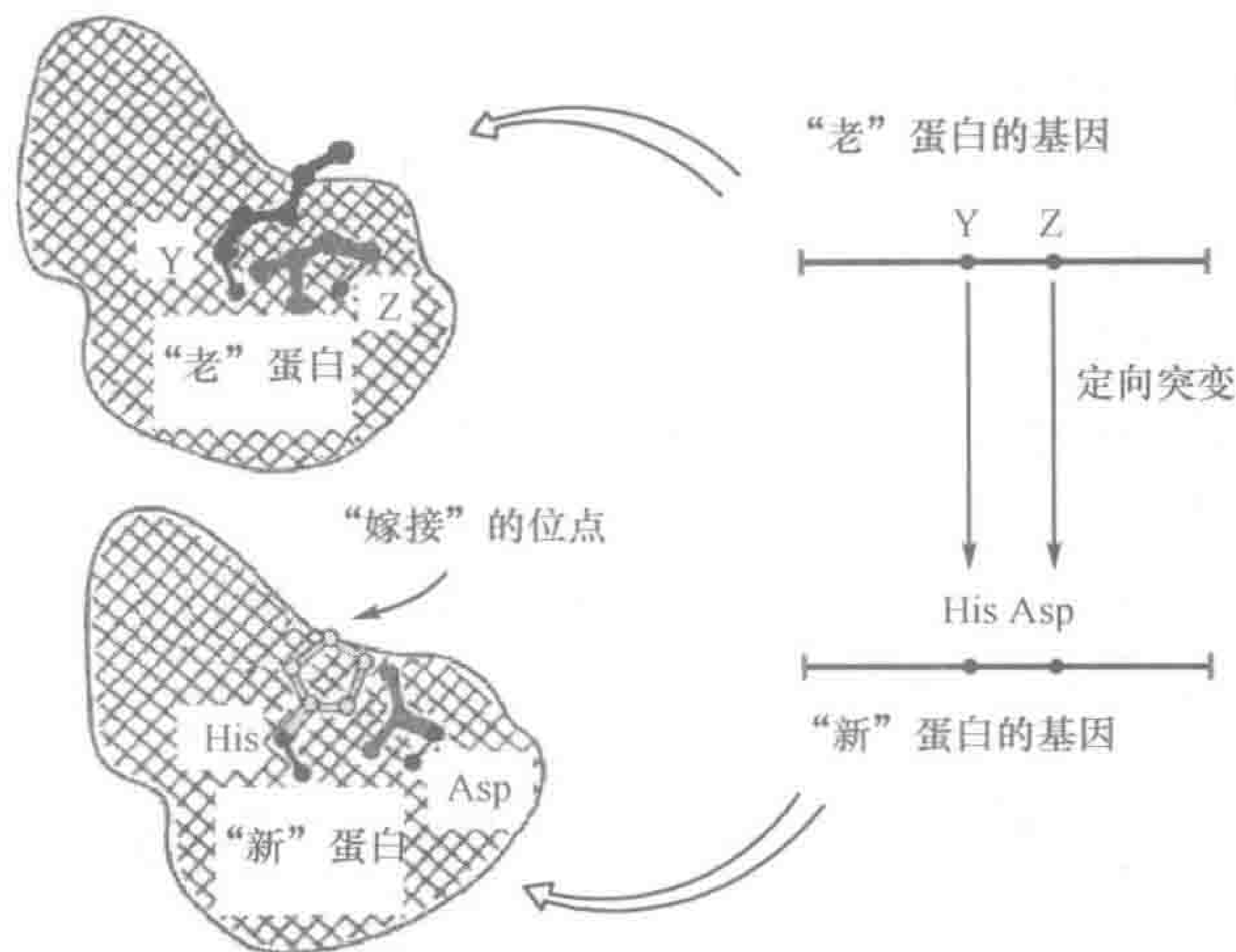


图 23-6　采用蛋白质工程向蛋白质"嫁接"新功能位点的示意图。图片取自 Finkelstein. 1989. Biopolymer and Cell. 5:89-93,有修改

■　我在图 23-6 中展示的只是一个简单的教学示例。而下面是 David Baker 研究组的一个杰出工作,他们成功地为一个不存在于生物体中的化学反应创造了一个在自然界中不存在的活性中心。实质上该活性中心取自"Kemp 消除反应":从一个芳香分子的突出部的质子转移。从消除反应过程的物理学和化学出发,完成了对功能基团的类型和构象的选择,这意味着使反应路径中的过渡态稳定化(顺便说一句,与其他几个一样,上文说到的 Asp-His 二联体位于芳香侧链基团的口袋中的特定构象有益于该催化作用。)。接下来,选出了(借助于上面已提及的程序包 Rosetta)可以把所选的二联体按所需的周围环境和所需的构象嫁接于其上的蛋白质。然后利用基因工程的方法完成嫁接,蛋白质被表达出来,并证明所得的人工酶把 Kemp 消除反应相对于其自发进行速度加速到 10^6 倍(作为开始,这已经很不错了,尽管某些天然酶能加速到 10^{12}倍);而 X 射线结构分析表明,嫁接的活性中心看上去实际上正如理论设计所要求的那样。于是,事情就做成了:获得了具有设计好的、但之前不存在于自然界的活性的人工酶。

■　蛋白质工程是解决各种问题的方法。它也被用于解释"同一条氨基酸序列能

否编码两个完全不同的蛋白质链折叠形状”的问题。回答是:如果在氨基酸序列上精心操作的话,就有可能。

Orban 和 Bryan 选取了链长相同、但结构和功能不同、序列也不同(同源性为15%,实际上符合随机相似性)的两个蛋白质:PSD1 和 GB1。接下来,逐步向这两个序列中引入彼此相近的突变,并且在每一步中控制让蛋白质保存它们最初的功能。Orban 和 Bryan 达到了让序列在维持自身结构(Alexander,et al. 2007. Proc Natl Acad USA. 104:11963-11968)和功能(He, et al. 2008. Proc Natl Acad USA. 105:14412-14417)的同时,使序列匹配达到 88%,即它们的 58 个残基中有 51 个相同(图 23-7)。在这两个序列上的后续工作进一步减小了它们的差别(而它们的结构和功能之间的差别保持不变)到只有 1 个残基不同(Alexander, He, Chen, et al. 2009. Proc Natl Acad USA. 106:21149-21154)!

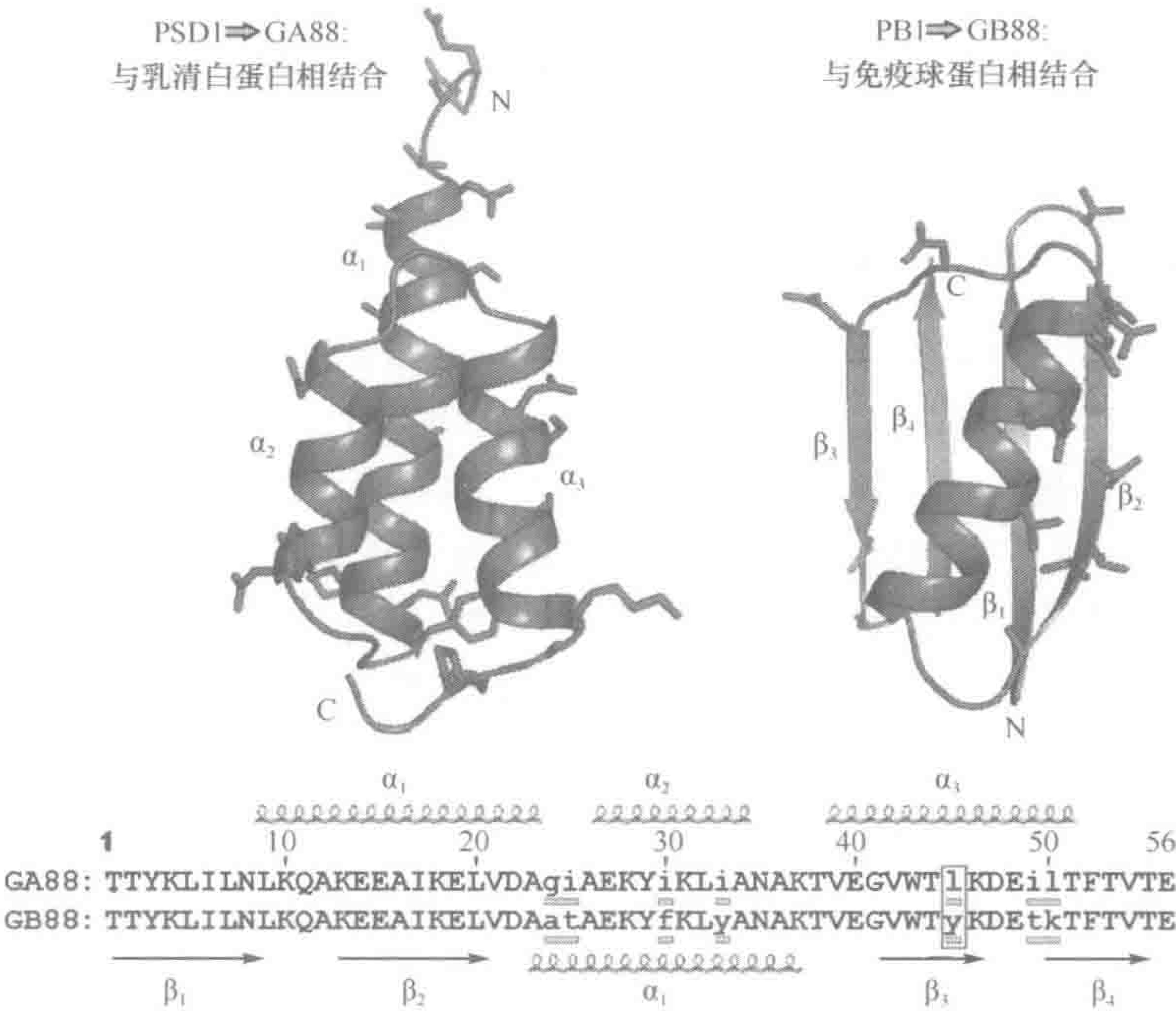


图 23-7 两个蛋白质在引入大量突变之后,仍然保留自身的功能和空间结构不变,但取得了 88%的序列一致性。图上方为蛋白质折叠形状,改变后的残基着彩色并显示了侧链基团的骨架模型。图下方显示的序列中,相同的氨基酸残基以大写字母表示,不同的以小写字母(加下划线)表示。方框括出的是这两个蛋白质在维持自身结构和功能的极限时有区别的唯一的残基。图片取自 He,et al. 2008. Proc Natl Acad USA. 105:14412-14417,略有简化和增补

在另一个有趣的工作中(Gambin, Schug, Lemke, et al. 2009. Proc Natl Acad USA. 106:10153-10158)证明,具有四螺旋束外观的蛋白质,在有目标性地引入一些突变之后,获得了处于另一种构象的可能性:也具有四螺旋束的外观,但螺旋折叠形状不同。而且一种环境条件能使一种结构稳定,另一种环境条件使另一种结构

稳定，而在中间条件下两种结构能同时共存。

因此，几乎相同或甚至完全相同的（但精心挑选的！）氨基酸序列在原则上可以编码两个完全不同的蛋白质链折叠形状。

■ 差不多在掌握了在"天然"蛋白质中引入突变之后，蛋白质工程就转向蛋白质分子的设计和结构设计。

结构设计是结构预测的逆向问题。如果在预测中我们需要找到对所研究序列最有利的空间结构，则在设计中我们需要找到对建立所希望的空间结构有利的序列。

总的来说，人造结构的计算应该比"天然"蛋白质的结构预测更简单（"人造"塔的刚度计算比"天然"树木要简单：它是如此设计的，以便于能进行计算！）。而新蛋白质的结构设计是以蛋白质结构理论为基础的；通常是以编码强有力的、内部稳定并能有效粘连的 α 和 β 片段的序列作为它们的"建筑模块"。

蛋白质设计是在 20 世纪 70 年代末期至 80 年代初期出现了创建人造基因的技术和精确的蛋白质链化学合成技术的时候建立起来的。在 80 年代末期至 90 年代初期，通过估算和试错法创造了最简单的蛋白质分子。它们的构造起始于天然蛋白质，而它们的氨基酸序列是如此选择，以致没有能使这些构造稳定的同源天然序列。

■ 第一个制造出来的（在 Degrado 研究组中）是四螺旋束（图 23-8）。设计是在与实验密切对话中进行的。对该蛋白质最终形式的研究表明，它是螺旋的和球状的；而且这个人造蛋白的结构显得比任何"正常"蛋白质对温度变化都要稳定、牢固得多。然后，证明了这个蛋白质确实在温度升高时不会熔化，因为从开始时它就是个熔球……后来通过向它里面引入能与离子结合的组氨酸（His），加强了这个人造蛋白质的刚性结构。而且这个蛋白质复合体已经显得与天然蛋白质一样是固态的。

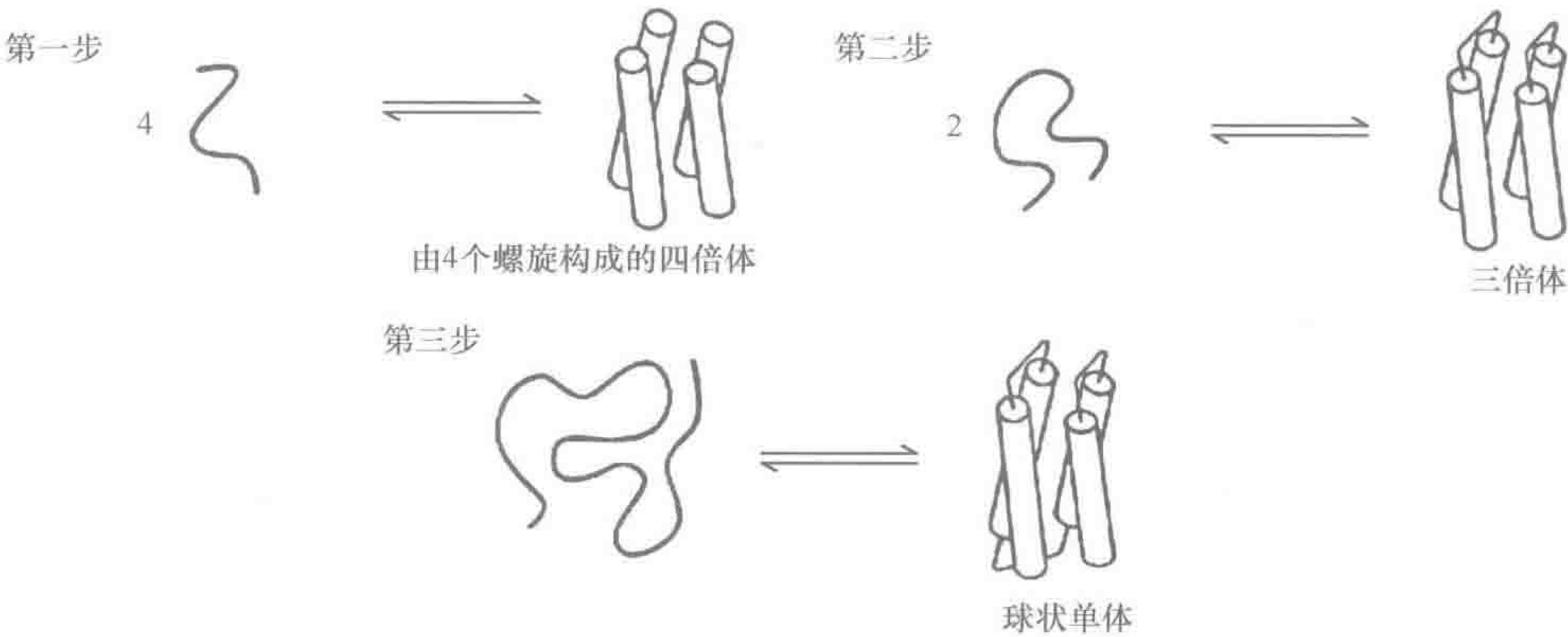


图 23-8 DeGrado 研究组完成的四螺旋束设计的主要步骤。第一步选择能粘连成四倍体的短螺旋。第二步设计能把这些多肽成对地连接起来的环线，并挑选出连接后能形成二聚体的那些种类。第三步设计最后的环线，并挑选单体四螺旋人造蛋白质。图片取自［6］，承蒙 W. F. DeGrado 的慷慨许可

在很长一段时间里,所有人造蛋白质(除了那些通过形成离子复合体而得以加强的之外)构成了一个极好熔球的集合。它们都具有非常好的二级结构,它们都非常紧密,但它们都没有足够的硬度。

■ 问题:为什么想要制造正常的“硬”蛋白质,却得到了熔球?

看来原因如下。因为大家都知道如何制作稳定的二级结构(Leu 和 Ala 在中间,Glu 在 N 端,Lys 在 C 端,就得到 α 螺旋,多一些 Val、Ile 和 Thr,就得到 β 串);大家也都知道,如何迫使这些 α 和 β 结构粘联起来:应该在它们上面用疏水基团制作连续表面(图 23-9);也知道如何迫使这些蛋白质不要聚集起来:应该用亲水基团制作这些 α 螺旋和 β 结构的对侧表面。但是至今还没有人知道如何在蛋白质疏水核中生成侧链基团的紧凑拼装的秘方。正是这种紧凑拼装包还没有人获得。而在能粘联、但不是紧密粘联起来的二级结构中如果没有紧凑拼装,获得的就只是熔球……

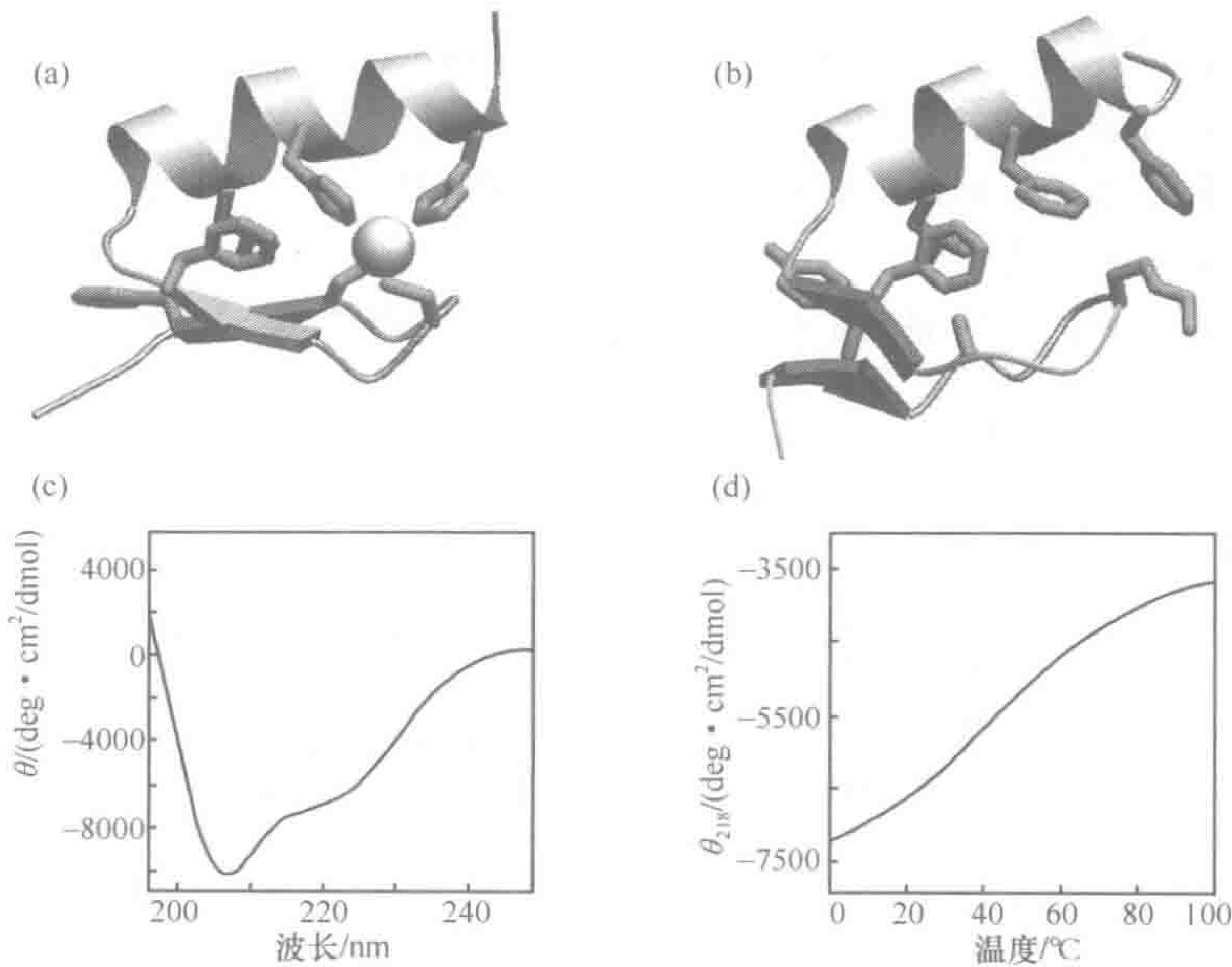


图 23-9 (a)原始的“锌指”结构(Zif268 蛋白的第二个模块);Zn 离子显示为小球。(b)人造蛋白 FSD-1 的结构。(c)FSD-1 在 1℃下的圆二色性(CD)谱。(d)FSD-1 随温度的圆二色性谱变化。图片取自 Dahiyat B. I. ,Mayo S. L. 1997. Science. 278:82-86,略有简化

■ 为了克服计算方法和设计方法中的精度不足,常常从“理性的”结构设计出发,并采取向链中引入大量随机突变(点状的或块状的,但一定是大量的!),随后挑选具有“蛋白质样”活性(如专一性地与某种东西相结合)的变体。通过这种途径,N. Wrighton 研究组成功地制造出了具有促红细胞生成素(erythropoietin)活性的迷你蛋白(由两个 β 发夹形成的二倍体)(有趣的是,这种天然激素由 166 个残基

组成,而它的人造类似物,总共只要 20 个!)。

■ 不过,看来,紧凑拼装的理性设计问题可以得到解决。只要有好的计算机算法对天文数字的侧链基团的可能拼装包进行轮换,并剔除无望的变体,紧凑拼装的计算(对于小蛋白)就可能得以执行。

至少,Dahiyat 和 Mayo 这样做了并于 1997 年创造了固态的(没有任何额外的与离子形成复合体的作用)小蛋白(图 23-9)。该蛋白是基于"锌指"结构(广泛分布的与 DNA 相结合的模域)而创造出来的,但没有支撑天然锌指结构的 Zn 离子。Dahiyat 和 Mayo 的人造蛋白 FSD-1 具有与天然锌指较低的(20%)同源性,并剥去了锌结合中心。虽然如此,它在低温下是固态的(使用核磁共振对它的结构进行了研究)。然而它的熔化温度范围很宽,即显然不像与它类似的天然蛋白质那么协同[图 23-9(d)]。

■ 除了基于天然构造的人造蛋白质之外,还有一些其构造没有天然类似物的人造蛋白质被设计制造出来了。

在我们的研究组由 A. N. Fedorov, D. A. Dolgikn 和 M. P. Kirpichnikov 设计并得到的人造蛋白质 albebetin(图 23-10)就具有一个至今在自然界中未观察到的结构。它是由两个 α-β-β 重复体(这就是它的名称来源)构成的,而且它不与形成球状蛋白结构的基本原则相矛盾。

对 albebetin 的结构研究表明,它具有指定的二级结构[图 23-11(a)],并且是紧密的,非常稳定,抗尿素的解折叠作用,抗空间结构的蛋白水解。然而它宁愿协同地熔化并处于熔球态,也不处于固态。

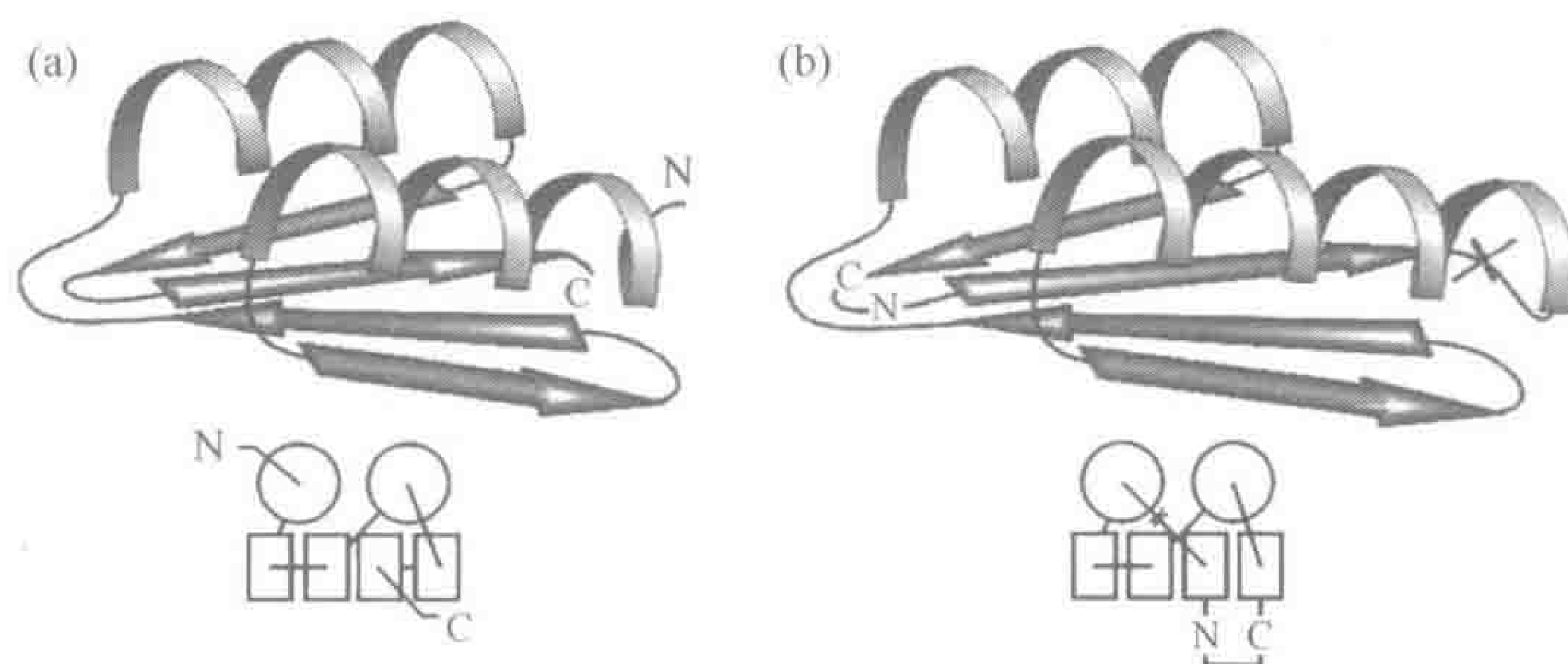


图 23-10 设计制造的 albebetin 的空间结构 (a)和实验测定的核糖体 S6 蛋白的空间结构示意图(b)。下方显示了这两个蛋白的拓扑示意图。对 S6 还显示了人工引入的置换突变:剪断环线(×)、连接原来的 N 端和 C 端(N–C)。这些置换突变使它具有 albebetin 的设计结构

通过另一种方法也获得一个具有为 albebetin 设计的结构的蛋白质:对天然蛋白质 S6[它和其他一些不久前才被解析出空间结构的蛋白质,具有为 albebetin 而设计的同样的结构片段折叠方式,但这些片段由蛋白质链以不同的方式连接,如图 23-10(b)所示]进行环形置换突变。所获得的蛋白质具有固态的、协同熔化的空间结构[图 23-11(b)]。

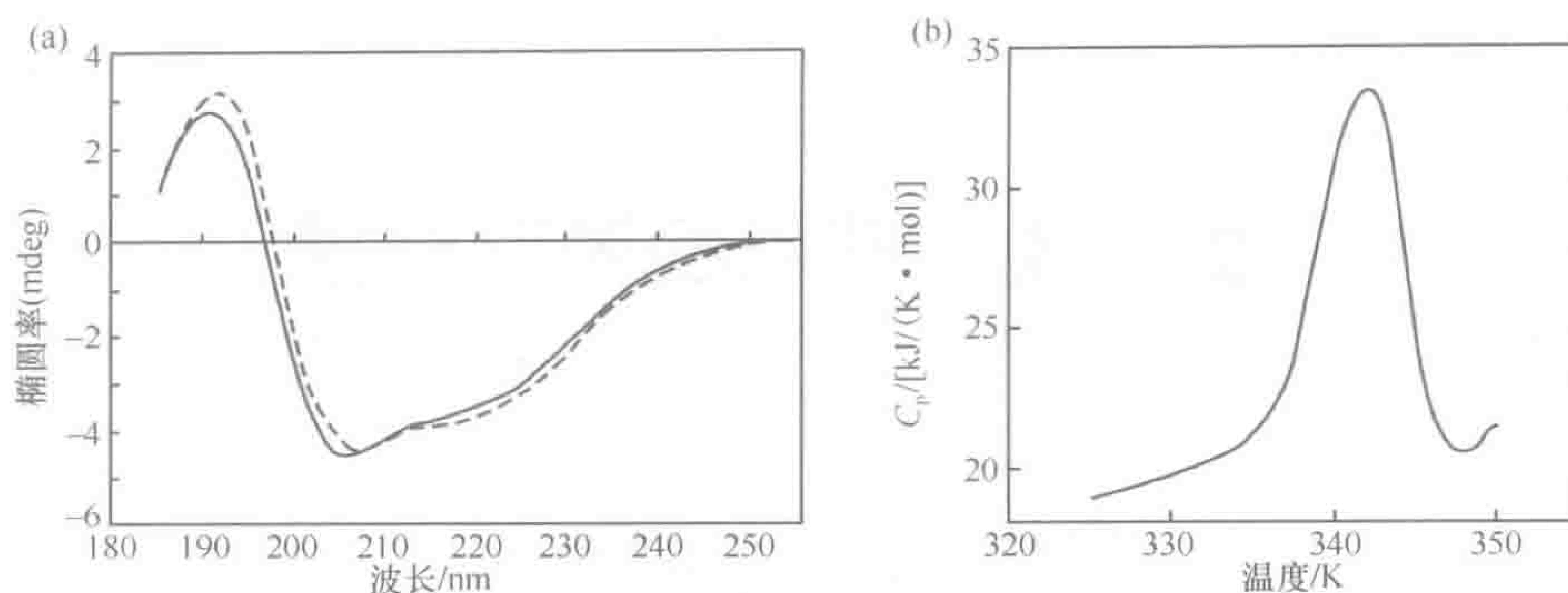


图 23-11 (a) albebetin (- - -)和 albeferon(——)的圆二色性谱;(b)为使其具有 albebetin 拓扑结构而进行了置换突变的蛋白质 S6 的微量热熔曲线。图片分别取自 Dolgikh D. A., Gabrielian A. E., Uversky V. N., et al. 1996. Appl Biochem, Biotech. 61:85-96 和 Abdullaev Z. Kh., Latypov R. F., Badretdinov A. Ya., et al. 1997. FEBS Letters. 414:243-246,略有修改

不久前,albebetin 被用作功能活性的载体。在它里面,更准确些,是在新的被称为 albeferon 的蛋白质中,被加入了人 α2 干扰素的 131-138 片段,该片段能激活胸腺细胞的母细胞转化。实际上,活性正是这个干扰素片段执行的,而这个蛋白质的其余的堆码成球状的肽链的用处是作为保护鞘,保护 131-138 片段不被蛋白水解,并使它不至于过于活跃地发挥作用。实验表明,albeferon 也正是这样工作的。

■ 对具有功能活性的人造蛋白质的研究工作正在许多研究组里进行。基于长螺旋束创造出了纤维蛋白质模型。由两亲性(疏水-亲水)螺旋创造出了第一个"能工作的"膜蛋白模型:这些螺旋形成离子通道,而且点突变能够剧烈地改变它们的选择性。通过使用化学性依附于血红素的 α 螺旋多肽还制造出了功能性的"蛋白质"。

■ 总而言之,在当今时代,蛋白质正在由令人崇敬和惊异的观测对象迅速地转变为活跃的工程活动目标对象(而惊奇还在继续……)。

蛋白质功能的物理学原理

第24讲(两倍)

蛋白质功能与结构。基本功能。结合蛋白:DNA结合蛋白,免疫球蛋白。酶。活性中心:球形酶功能的“缺陷”。蛋白质的硬度对基本酶功能很重要。催化和底物结合中心。抑制作用。辅因子。多价离子。酶催化的机制。例子:丝氨酸蛋白酶。催化中的过渡态理论及蛋白质工程方法对其的证明。抗体酶。催化作用的特异性。“钥匙-锁”概念。

本讲我将专门讨论蛋白质的功能。

关于这方面可以讲很多。但是我只谈谈蛋白质工作的物理基础。我的讲解将只给你们来自功能蛋白质生活中的一些图片,它们强调的是蛋白质的空间排列对其功能的关键作用。一些这类的图片已经在前面的章节中(尤其是在讲述膜蛋白时)出现过了。所以今天我将只介绍水溶性球状蛋白。

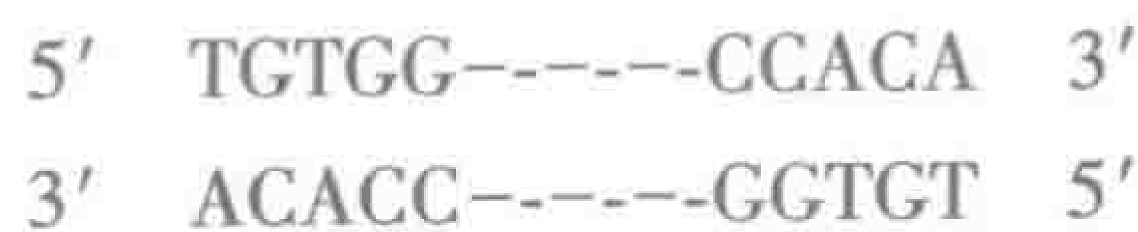
■ 蛋白质行使功能的基本过程可以非常粗略地表示如下:

结合(bind) → 变换(transform) → 释放(release)

准确地说:有些蛋白质只执行其中的部分行为;“结合”和“释放”可能指的是结合和释放几个不同的分子;而“变换”可能意味着化学转变、构象变化(蛋白质本身以及底物)和/或在空间中的移动(蛋白质或底物)。

■ 让我们从基本功能为“结合”的蛋白质开始吧。例如DNA结合蛋白就属于这一类。

为了与DNA结合,蛋白质表面,而且是很大一部分,应该与双螺旋表面近似地互补[图24-1(a)]。这样蛋白质表面的突出部可以深入到DNA的斜沟中,在那里蛋白质侧链基团可以对具体的DNA序列进行精确识别(图24-2)并与此蛋白质预定的DNA序列相结合。图24-1中所显示的蛋白质都是二倍体,而且正是在这种形式下它们能与DNA双链体互补。这种二倍体的两个相同的可识别DNA的α螺旋可以识别DNA双螺旋中的**回文序列**,即DNA双链体中在绕垂直于DNA轴旋转180°后序列与原来相同的片段,例如:



其中,“---”表示回文序列左右两半中的任意 DNA 序列,而“ · ”表示旋转轴。

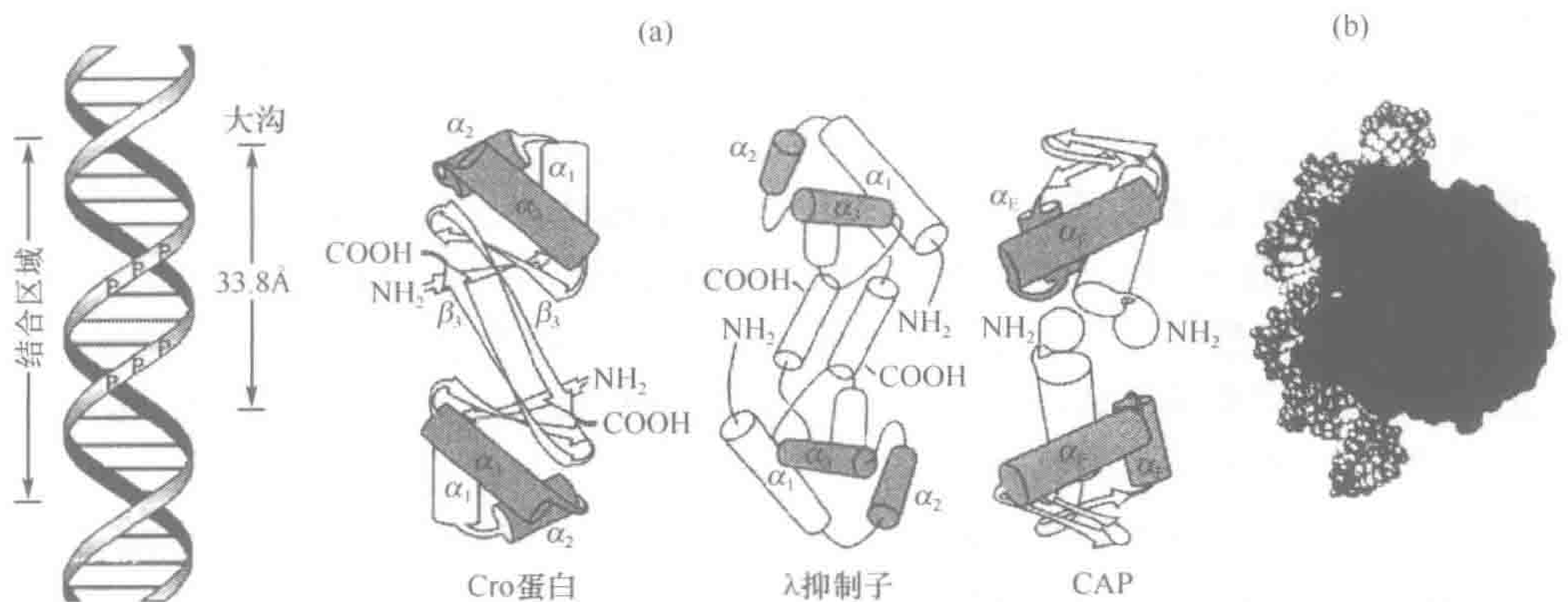


图 24-1 (a) DNA 结构(左)和几个带有典型的 DNA 结合模域“螺旋-转角-螺旋”(灰色所示)的蛋白质。分解代谢物基因激活蛋白(CAP)只显示了其 C 端域。所有这些蛋白质都是二倍体,都能以其 α_3 螺旋(CAP 的 α_F)识别 DNA 的斜沟。二倍体中的这些螺旋间的距离很接近于 DNA 的双螺旋周期(33.8Å)。图片经 B. W. Matthews 许可取自文献[6]。(b) 与二联体 CAP 蛋白(右边黑色)结合的 DNA 的弯曲(左边明亮的螺旋)。CAP 蛋白与 DNA 的结合需要环腺苷酸(cAMP)的存在

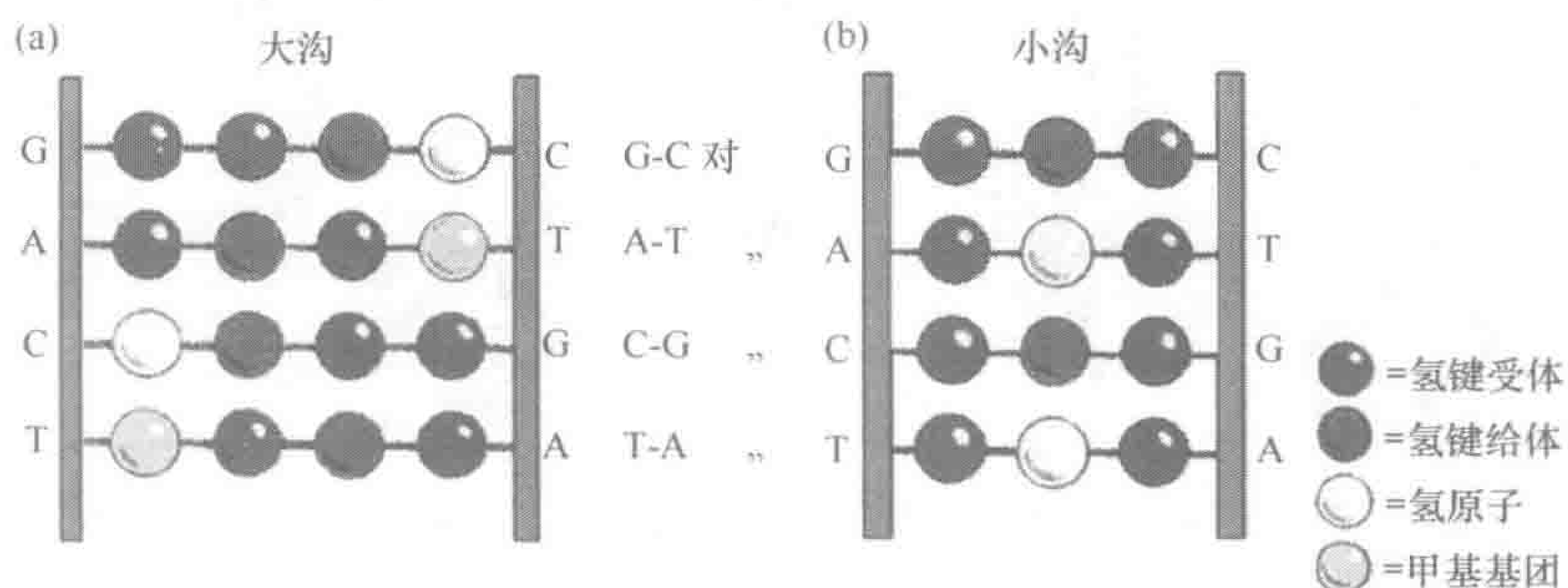


图 24-2 在 DNA 大沟和小沟中 A-T 和 C-G 碱基对的不同功能基团所形成的特征模式。图片经许可取自文献[5]

在这种蛋白质二倍体中,与 DNA 结合的 α 螺旋是彼此**反**平行的,而且它们之间的距离接近于 DNA 双链的周期长度,因此二倍体可以嵌合到 DNA 双链的一侧。但是,不同的蛋白质的这些 α 螺旋相对于结合中心轴的倾斜角不同,这就导致 DNA 在结合时产生不同的弯曲度。有时候这种由于结合而产生的弯曲相当大[图 24-1(b)]。

有时候蛋白质与 DNA 的结合需要**辅因子**的存在。它们能使蛋白质的结构发生改变,更准确地说,轻微地变形,并导致它从无活性状态转变为活性状态。例如,

在色氨酸阻遏蛋白(Trp-repressor)(在大肠杆菌中该蛋白阻遏负责编码色氨酸合成所必需的蛋白质的 RNA 合成的操纵子)中,这种**辅**因子(确切地说是**辅**阻遏物)是色氨酸本身。当色氨酸没有与该蛋白结合时,在色氨酸阻遏蛋白二倍体中的 DNA 结合螺旋之间的距离比所需的 34Å 要小(约 28Å),这使得 DNA 不能结合。色氨酸与蛋白质结合后,螺旋被移开,从而能与 DNA 双链的斜沟互补并与之结合。因此,当细胞中的色氨酸很多时,它与色氨酸阻遏蛋白相结合,阻止合成更多的色氨酸合成蛋白,从而阻止自身(色氨酸)合成更多。这种调节方式称为**负反馈**。

在这种情况下,色氨酸既是作为色氨酸阻遏蛋白的 DNA 结合活性的**刺激剂**起作用,同时是也色氨酸合成所必需的蛋白质合成的抑制剂。而且色氨酸阻遏蛋白的这种刺激是"**变构性的**(allosteric)",因为色氨酸与蛋白质"**在另一部位**"相结合,而不是在与 DNA 结合的部位。

图 24-1 和图 24-3 中所显示的"螺旋-转折-螺旋"模域(motif)是典型的,但远不是唯一的能结合 DNA 的结构模域。为了说明这一点,我在图 24-4 中给出另外三个典型的模域。我想强调的是,DNA 结合蛋白可以归属于不同的结构类型(在图中有 α 型和 α+β 型蛋白质),而且实际上与 DNA 结合的可以是 α 结构,也可以是 β 结构。

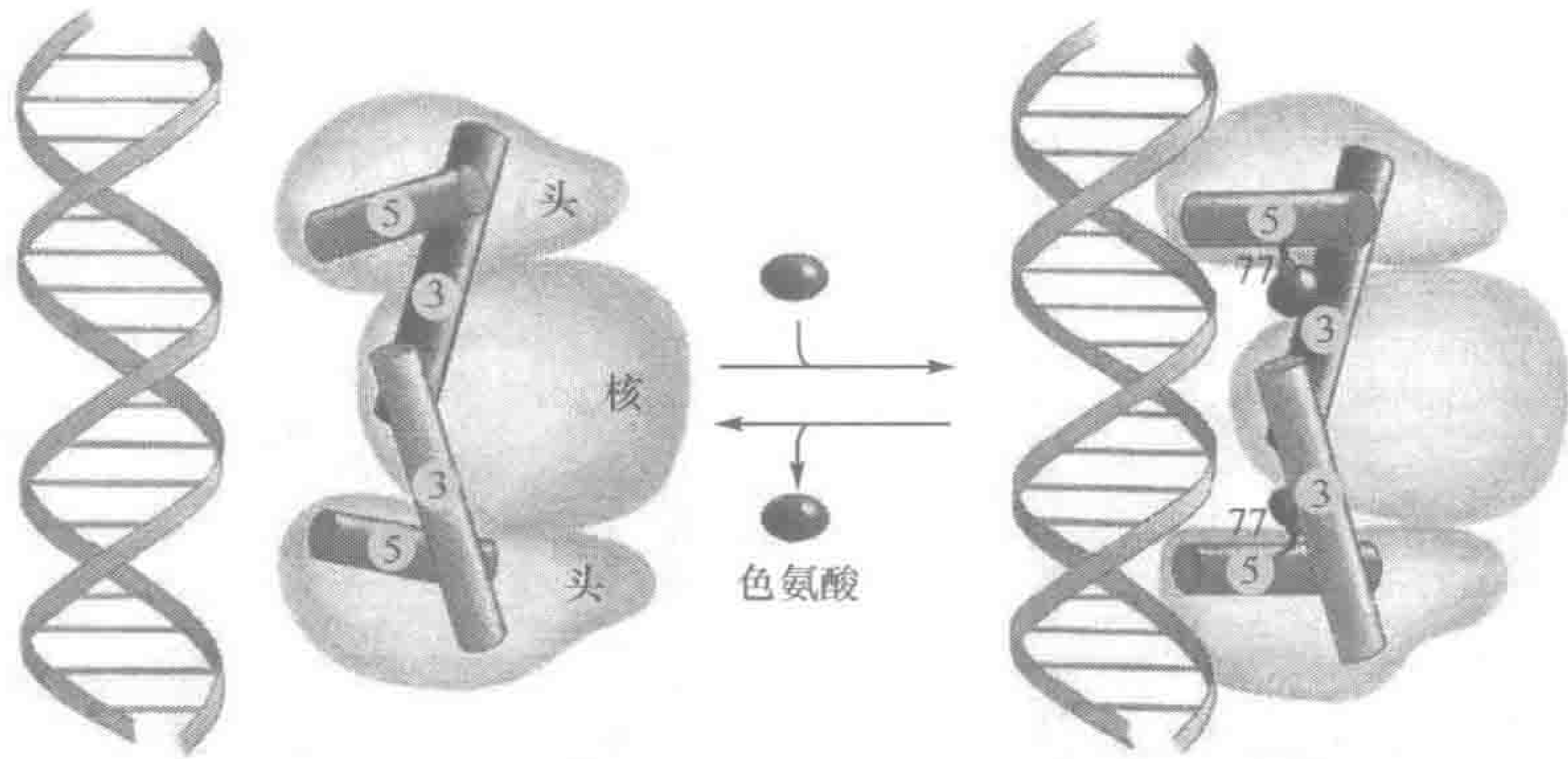


图 24-3　色氨酸(Trp)阻遏蛋白作用示意图。在含有一个共同融合核心和两个相同头部的二倍体的基本轮廓背景上,只显示了核心与头部之间的两个螺旋(3 和 5),其间可以嵌入辅阻遏物(色氨酸)。辅阻遏物推开 77 号残基,从而使 α 螺旋 5 偏移。只有在这种情况下,α 螺旋 5 才能与 DNA 相结合。图片经许可取自文献[5]

到此为止我们已经讲述了关于能与 DNA 斜沟相匹配的蛋白质结构(特征尺寸:10~30Å)的大致特征。对应当与蛋白质相结合的特定 DNA 序列的识别,是由蛋白质表面(特征尺寸为原子尺寸,约 3Å)的更精细性质负责的。

尽管对每一个已解析的 DNA-蛋白质接触的细节进行观察能让我们准确地看到在蛋白质和核酸侧链基团之间有哪些氢键,以及在这些特定位点促成 DNA 与蛋白质相接触的其他近距离联系,遗憾的是,对蛋白质用于对 DNA 片段进行选择性识别的"通用编码"仍然不清楚(甚至不清楚,是否存在这样的"编码")。

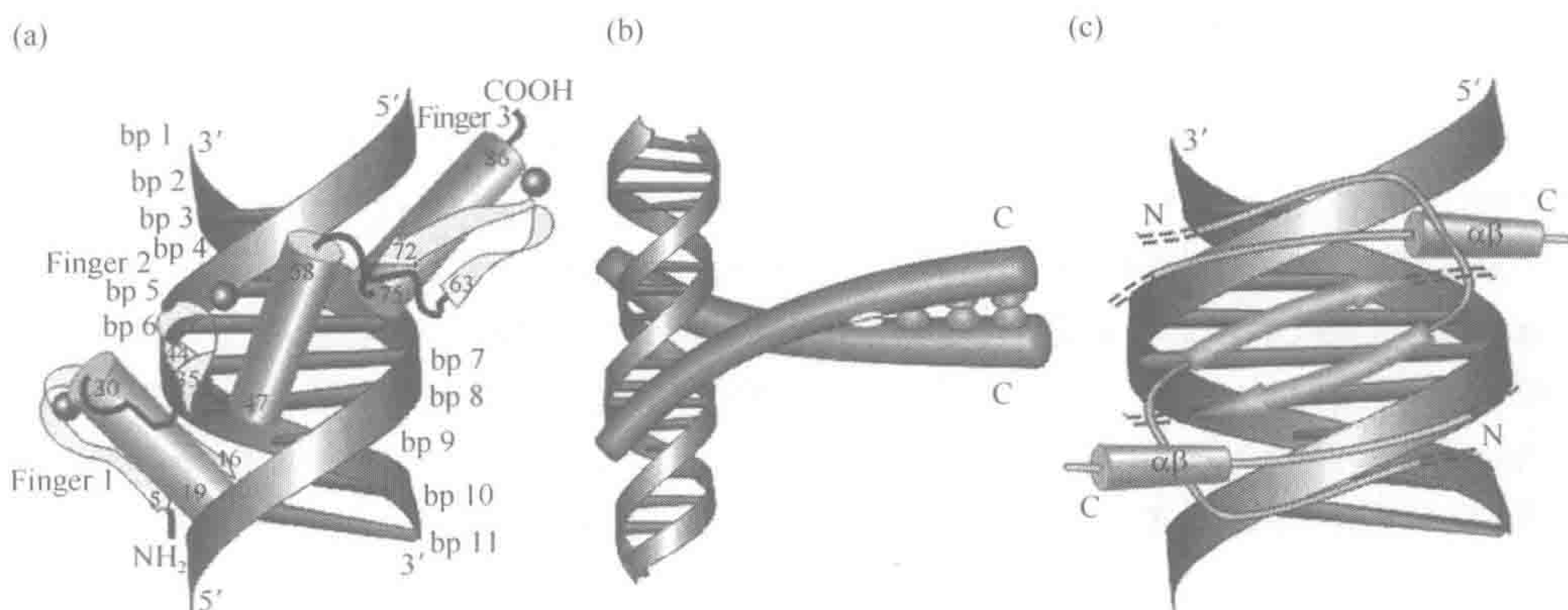


图 24-4　其他三种典型的 DNA 结合蛋白模域。在其中的两种中，是 α 螺旋起关键作用：(a)“锌指”(图中小球为锌离子)和(b)“亮氨酸拉链”。在第三种(蛋氨酸阻遏蛋白)中起关键作用的是 β 发夹(c)：其中的 β 发夹特异性地与 DNA 的斜沟相结合，同时 α 螺旋 αB **非**特异性地与 DNA 的糖-磷酸骨架相结合。在已知的球状蛋白中，锌指(这种域可以从整个蛋白质上截断并单独分离)是最小的，而亮氨酸拉链是结构上最简单的。当亮氨酸拉链未与 DNA 结合时，它只是由平行 α 螺旋组成的二聚体，狭窄的疏水表面沿其全长粘在一起。形成这些表面的侧链基团显示为突起。但是，每一个螺旋有一个**非**疏水性的侧链基团。在螺旋交叉点有插入在疏水表面的极性天冬酰胺(Asn))。它对形成此二聚体是必不可少的。因为它被更疏水的残基替代后可导致组装成不是二链而是三链或四链的复合体。图片取自文献[6](a)和文献[5](b)，(c)，略有修改。各链中显示了 N 端和/或 C 端

■　蛋白质对其他分子的高度选择性识别可以清楚地用**免疫球蛋白**(immunoglobulin)，或**抗体**(antibody)为例说明。这类蛋白质的任务(在哺乳动物体内)是对不同大小不同分子中的小尺寸(一个原子或几个原子的尺寸)**抗原决定簇**(antigenic determinants)①进行精确识别。而与免疫球蛋白相似的 T **细胞受体**以同样方式识别特定细胞(如病毒感染的细胞)中的小抗原决定簇。

免疫球蛋白是由许多 β 结构的域以及域之间的相对不大的柔性铰链组成的[图 24-5(a)]。可变域(抗原结合域)组合的多样性保证了免疫球蛋白种类的广谱性，相应地，保证了它们活性的广谱性。而这些域的刚性保证了每一种免疫球蛋白

---

①抗原决定簇(antigenic determinant)或称抗原表位(antigenic epitope)是位于抗原物质分子表面或者其他部位的具有一定组成和结构的特殊化学基团，它能与免疫系统中淋巴细胞上的受体及相应的抗体分子结合，它是免疫原引起机体特异性免疫应答和免疫原与抗体特异性反应的基本构成单位。抗原决定簇决定着抗原的特异性，即决定着抗原与抗体发生特异结合的能力。一般来讲，蛋白质抗原的 3 ~ 8 个氨基酸残基可以构成一个抗原决定簇，多糖抗原中的 3~6 个呋喃环可以构成一个抗原决定簇。抗原分子中直接由分子基团的一级结构序列决定的决定簇称为顺序决定簇(sequential determinant)。抗原分子中由分子间特定的空间构象决定的决定簇称为构象决定簇(conformational determinant)。对蛋白质抗原而言，顺序决定簇决定于肽链的氨基酸序列即肽链的一级结构，而构象决定簇决定于由这个一级结构折叠而成的空间构象。

的活动都具有高度选择性。克隆选择理论①解释了免疫球蛋白多样性的起源，我不想在这里详述这一理论的基本原理。

[你们应该从其他课程中知道，胚胎细胞并**不**包含免疫球蛋白的轻链和重链的完整基因，而只包含它们的片段。在胚胎细胞基因组中，这些片段聚集成盒状排列：各种类型的重链可变域的三个片段单独排列，轻链的可变域、所有链的不变域、铰链也分别单独排列。在免疫体细胞形成期间这些片段被多次洗牌，且它们的超可变部分以某种神秘的方式进行突变，并结合成为免疫球蛋白的轻链和重链的完整基因。至今为止仍不知道这些"体细胞突变"是如何引入的，以及它们是如何引导进入基因的特定部位。而且，虽然它们在域上可能达到上千个（一起引入并由"幸存的"蛋白随后进行选择实在是太多了！），但是为什么这些突变没有使蛋白质结构爆裂，也是一个谜。]

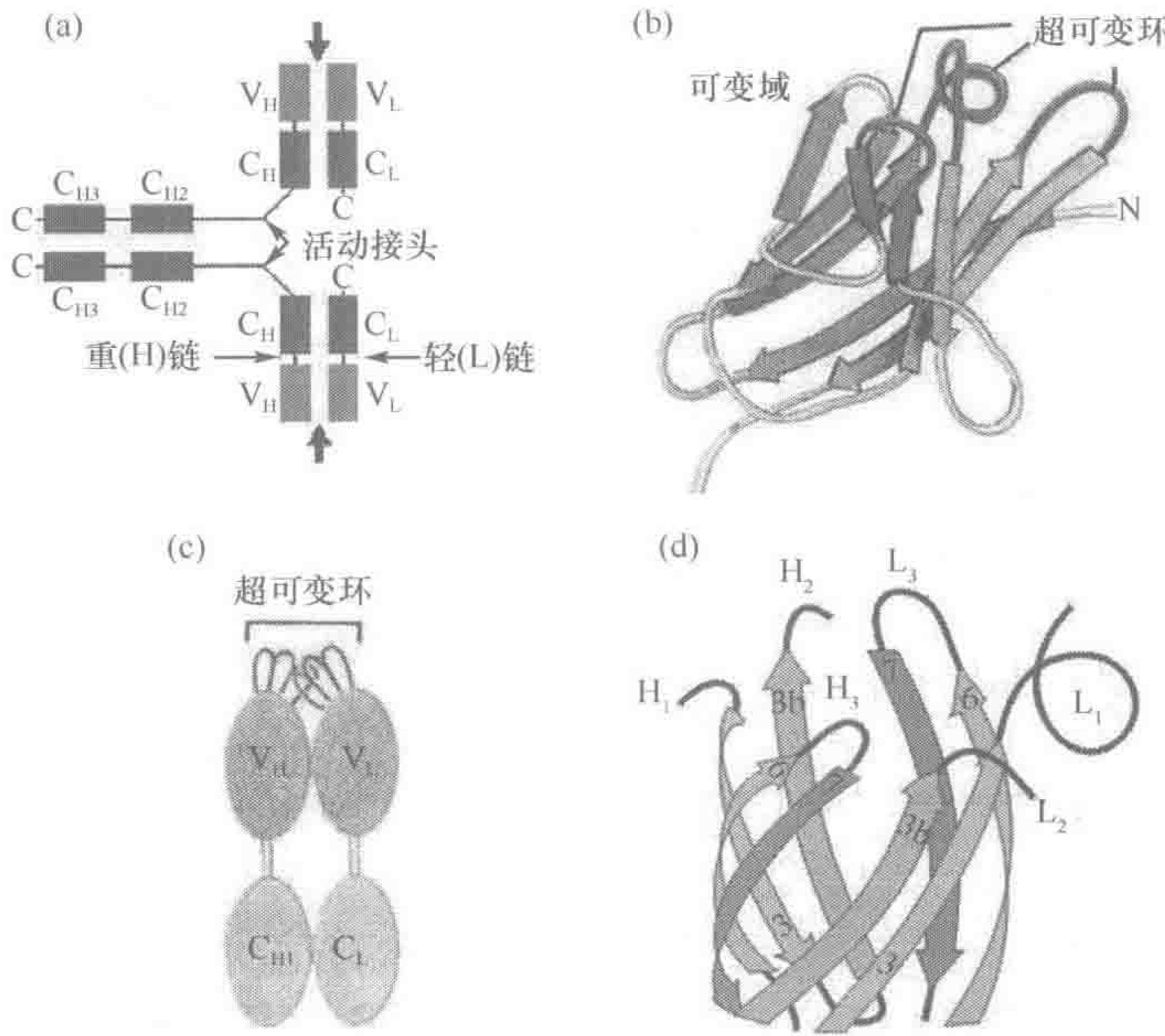


图 24-5 (a)一个免疫球蛋白（IgG）的总体结构。标出了两个轻链（L：各有两个域）和两个重链（H：各有四个域）的可变域（V）和不变域（C）以及所有链的 C 端。C_{H2}域是糖基化的。粗箭头指向在 V_L 域和 V_H 域之间的抗原结合"口袋"。(b) V_H 域的结构排列；V_L、C_L、C_H 和其他域的结构与之相似。在 V_H 域中显示了超可变环；V_L 域中也有超可变环。这两个域一起构成了抗原结合口袋(c)。(d)由 V_H 域和 V_L 域的 β 片构成的反平行 β 圆筒。抗原结合口袋是由重链和轻链的这些 β 片上伸出的环线（分别是 H_1，H_2，H_3 和 L_1，L_2，L_3）所构成的。图(b)、图(c)经许可取自文献[5]。图(d)取自文献[5]，略有修改

①抗体产生的克隆选择理论认为，动物体内存在着许多免疫活性细胞克隆，不同克隆的细胞具有不同的表面受体，能与相对应的抗原决定簇发生互补结合。一旦某种抗原进入体内，与相应克隆的受体发生结合后，便选择性地激活了这一克隆，使之扩增并产生大量抗体即免疫球蛋白，抗体分子的特异性与被选择的细胞的表面受体相同。克隆选择学说的核心论点是：带有各种受体的免疫活性细胞克隆早已存在，抗原的作用只是选择并激活相应的克隆；细胞受体和该细胞后代所分泌的产物（抗体）具有相同的特异性。

现在对我们重要的是,抗原是由轻链和重链的可变域(V_H 和 V_L)共同识别的,确切地说,是由位于这些域交界表面的抗原结合口袋边缘的超可变环识别的[图 24-5(b)~(d)]。这些环的一级结构随免疫球蛋白的不同而不同(这导致了免疫球蛋白种类的极大多样性)。然而,对每一种类型,无论是氨基酸序列还是所有环线的构象都是严格固定的。而抗原结合口袋恰好位于由可变域的紧密连接的反平行 β 片形成的硬 β 圆筒上。所以每种免疫球蛋白分子都只能与特定的抗原决定簇紧密结合,而不与其他的相结合。

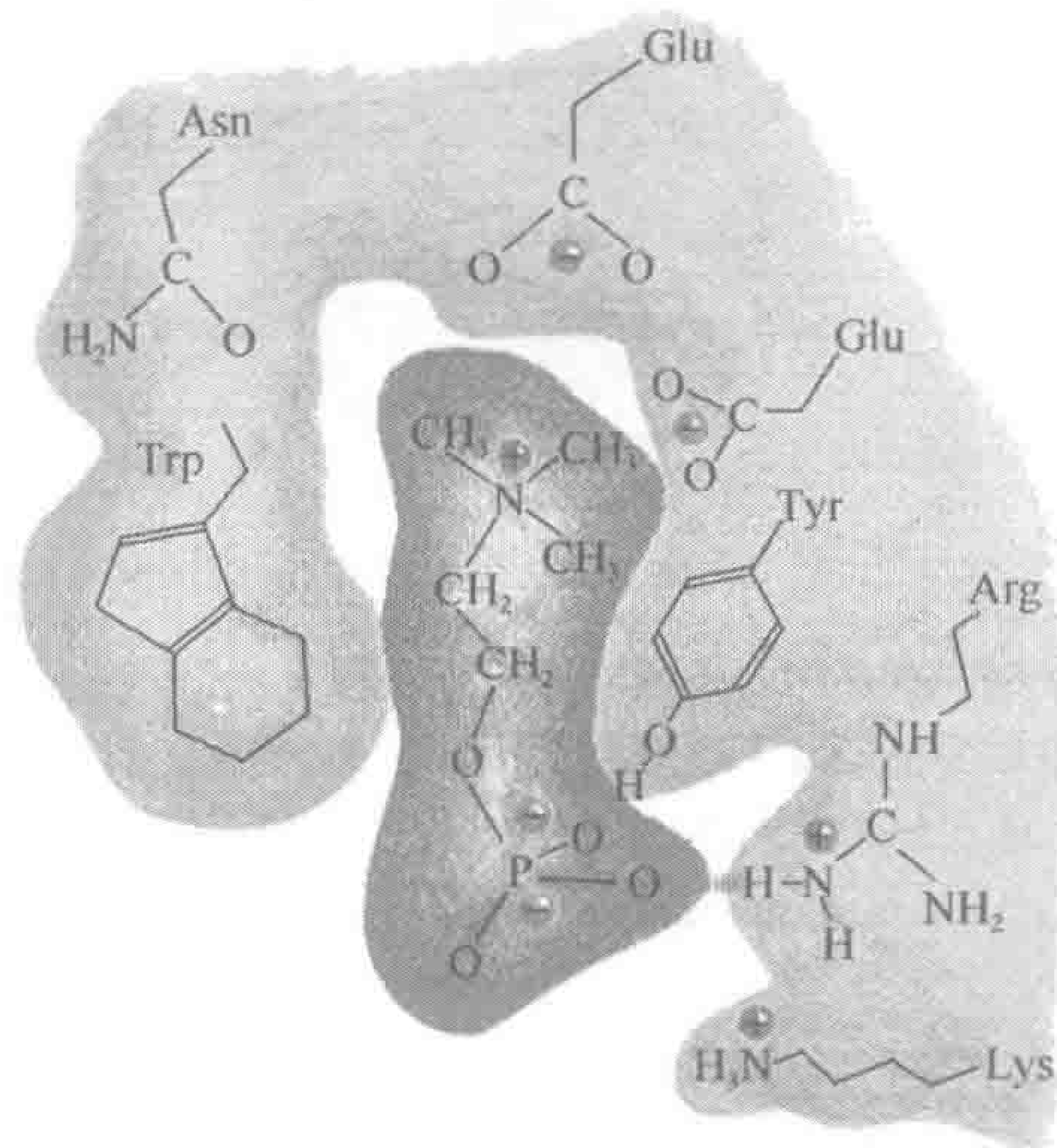


图 24-6 抗原和与之相结合的抗体"口袋"的特异性相互作用。图中显示了邻近的电荷和形成的氢键。图片经许可取自文献[5]

图 24-6 显示,抗原决定簇的结合选择性**不是**由蛋白质**整体**结构所决定的(它只是作为一个基座),而是由待结合分子的形状与相对较小(与整个域相比)的抗原结合口袋的互补性所决定的。

除此之外,被结合分子的疏水部分与口袋的疏水部分相接触,它的电荷与包埋在口袋中的电荷互补,而且抗原的氢键受体或供体与包埋于抗体口袋中的氢键供体或受体互补。所有这些使得结合(但只是严格地与特定抗原的结合)变得牢固。

■ 在很多其他与免疫球蛋白毫不相关的蛋白质中也观察到了与在抗体中相同的活性中心位置:在 β 圆筒上部的喇叭口处。例如,这也是在 α/β 圆筒(与免疫球蛋白不同的是,β 结构是平行的)中的活性中心的标准位置。通常,在研究蛋白质结构时容易注意到,活性中心常常位于它们的"标准缺陷"处,在标准地分布在蛋白球结构中的凹陷位置上(图 24-7):这种凹陷(由氨基酸链折叠模式所决定,而非由侧链基团决定)自动地使得有许多蛋白质侧链同时包围底物。

域接缝处也是常见的活性中心位置。图 24-8 中显示了胰蛋白酶型丝氨酸蛋白酶的活性中心,它位于两个 β 结构的域的接合界面上。

■ 在结束对蛋白质**结合**功能的讨论时,应该对它的对称功能"**释放**"说几句。

需要提及的是,许多被称为"天然解折叠"的蛋白质只有在与配体(或其他蛋白质、DNA 或 RNA)结合之后才具有其唯一的空间结构:要知道,只有在蛋白质与配体结合的时候,蛋白质用于保持结合特异性的结构刚性才是必需的。因此并不惊奇的是,在执行"释放"功能时,一些蛋白质会失去自己的唯一空间结构。例如,结合视黄醇的蛋白质在释放视黄醇之后转变为熔球。

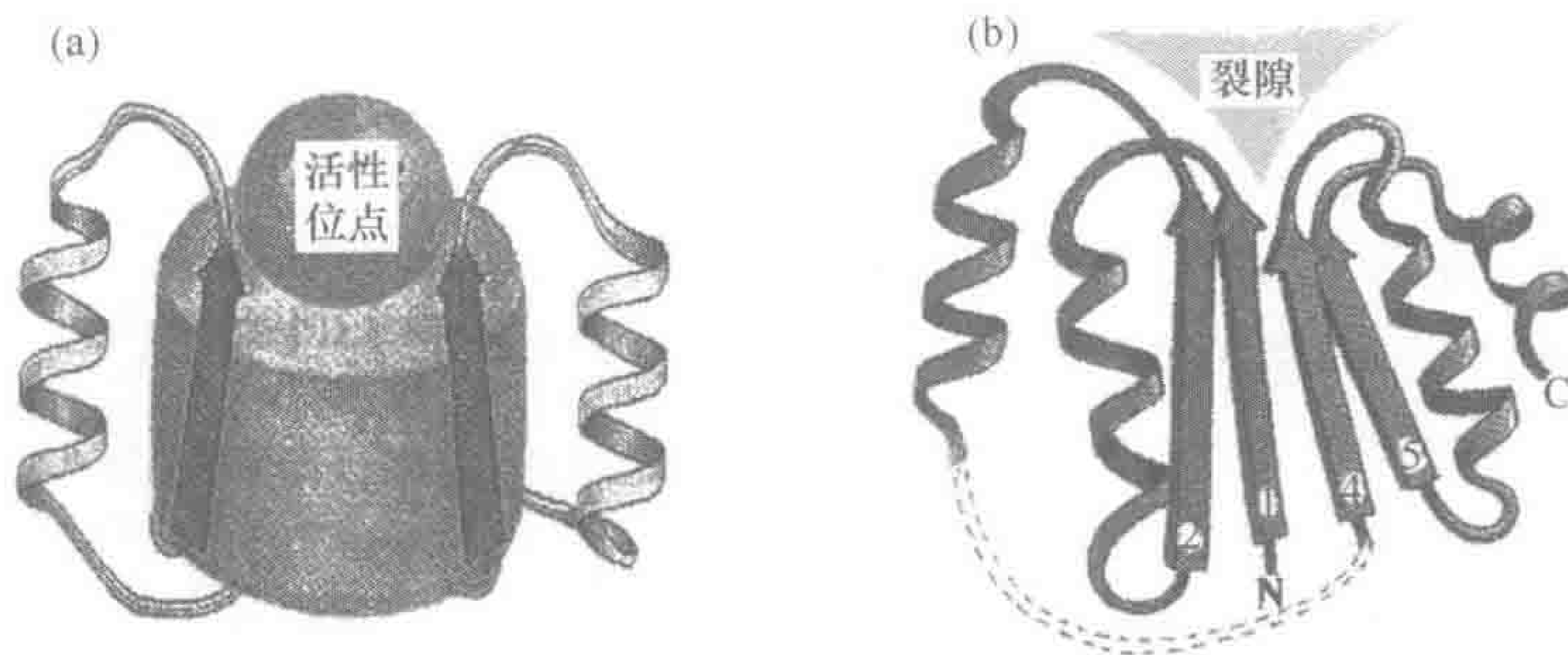


图 24-7　蛋白质球体结构上的标准凹陷常常决定活性中心的位置(不是其功能!)。(a)活性位点在平行 β 圆筒的 β/α 圆桶上方的喇叭口。在**反**平行 β 圆筒上方喇叭口的类似的活性中心位置参见图 24-5(d)。(b) Rossmann 折叠裂隙处的活性中心,位于右边的走向方向不同的 β-α-β 超螺旋的岔开位置(在超螺旋 β_1-α-β_2 中链远离我们,在超螺旋 β_4-α-β_5 中链走近我们)。图片经许可取自文献[5]

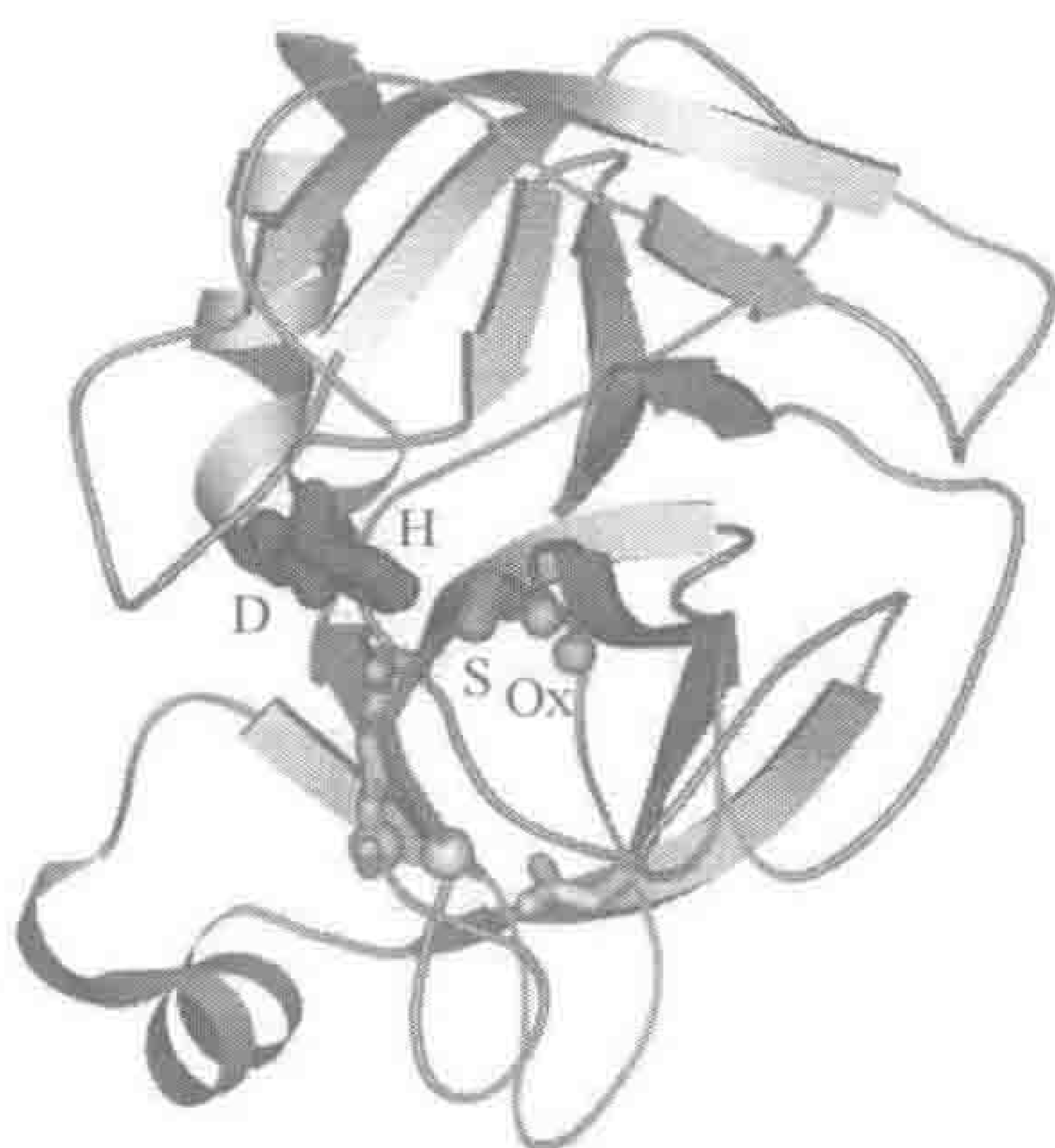


图 24-8　胰蛋白酶型丝氨酸蛋白酶的活性中心位置。图中显示了活性中心部分:催化中心,该处灰色表示"电荷传递三人组"[Ser195(S),His57(H)和 Asp102(D)]的侧链基团,以及底物结合中心,该处灰色小球表示组成氧阴离子洞(Ox)的 NH 基团,黑色表示非特异性的底物结合区,亮白色表示平铺于特异性底物结合口袋的基团。域接缝位于氨基酸对(His57,Asp102)和活性中心其他部分之间

■　现在自然应该讨论酶了。酶这类蛋白质的主要功能是对所结合的分子进行化学性转变。

酶既不能创造新的反应也不能改变反应方向,它们"只能"加速自发进行的过程。但是有时能加速几千亿(10^{12})倍,比最强效的催化剂都要更强几百万(10^6)倍。

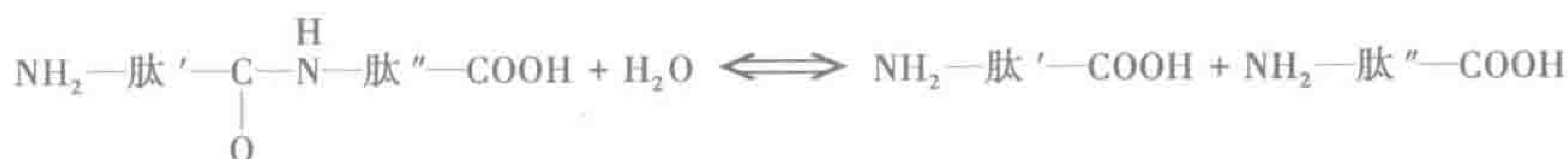
在此我不想介绍酶的分类(你们可以在任何一本生物化学教科书中找到①),

①国际酶学委员会规定,按酶促反应的性质,可把酶分成六大类。A. 氧化还原酶类:催化底物进行氧化还原反应;B. 转移酶类:催化底物之间进行某些基团的转移或交换;C. 水解酶类:催化底物发生水解反应。D. 裂解酶类:催化一个底物分解为两个化合物或两个化合物合成为一个化合物;E. 异构酶类:催化各种同分异构体之间相互转化;F. 合成酶类:指催化两分子底物合成为一分子化合物,同时还必须偶联有 ATP 的磷酸键断裂。

而专注于讨论酶是如何实现如此强烈地加快化学反应的问题。

丝氨酸蛋白酶(serine protease)是在讲述简单酶促反应时所使用的经典例子,我也不想违背这个传统。

丝氨酸蛋白酶切割多肽链,即进行如下反应



当环境中有足够多的水时,肽链的水解反应可以进行。但是进行得非常**缓慢**,需要很多年才行。换言之,水解反应(存在游离水时)是热力学有利的,但是需要克服很高的活化能垒。如果在环境中无水,则反应朝另一个方向进行,即合成肽链并释放水,但是反应速度也非常慢。

在有酶存在的情形下,肽链的水解反应(或者,在缺少自由水的情形下,由小片段进行肽合成的逆反应)只需几分之一秒,即酶急剧地降低了它的活化能垒。请看它是如何做到的。

首先来看一下酶活性中心的基本组成。

它由**催化反应中心**和**底物结合中心**组成,前者负责化学转变,而后者的任务是准确地把底物放置到催化切割刀下(或者说,放置于焊接/切割机器下,因为酶是同等地加速正反应和逆反应)。

在丝氨酸蛋白酶中,催化作用是由特定的丝氨酸(在所有胰蛋白酶家族中,它都被称为“Ser195”:根据它在胰凝乳蛋白酶中的位置。图 24-8 中显示的正是胰蛋白酶)的侧链直接完成的(参见彩图立体图以 D-31~D-34)。回忆一下丝氨酸侧链基团的化学式:—CH_2—OH。然而,为了能催化水解反应,丝氨酸需要有所准备:当氧处在—OH 中时,它不具活性;当 H^+被移除,氧变成 O^-状态时它才被活化。把 H 原子从 Ser195 上移除的工作是由“电荷传递三人组”中的另外两个成员:His57(它接受 H^+原子)和辅助性的 Asp102 执行的。对这两个残基进行突变(更不用说对起催化反应的丝氨酸残基进行突变或化学性修饰)实际上可以消除丝氨酸蛋白酶的催化活性。

Ser195 侧链基团的活性 O 原子在水解中起关键作用。它攻击被分裂的肽基团的 C 原子,把它纳入到一个临时的共价键中(该键的键长比普通的 C—O 键要长,不那么坚固),从而把 C 原子转换为四面体结构形式(此时它不是与三个而是与四个原子相结合:附加的原子就是 Ser195 侧链基团的 O 原子)。此复合物随后解体,使 C-N 键断开,导致所处理的肽链断裂。

底物结合中心是由能与断裂肽基团的氧相结合的**氧离子洞**(oxyanion hole)和**非特异性**的肽结合**区**(该区与氧离子洞一起,负责把待断裂的肽基团固定到相对于 Ser195 侧链基团的活化 O 原子的正确位置)以及负责氨基酸(在该氨基酸的羧基之后产生肽链断裂)识别的**特异性**的底物结合**口袋**组成(图 24-8,图 24-9)。

反应的大体过程如图 24-9 所示。这个示意图是经过多个研究组的长期工作

后才获得的。为构建这个图而综合了多种数据：不同底物的催化分解数据、酶的化学性修饰数据、蛋白质工程研究成果（它确定了酶的催化参与位点）、酶与未断裂的相似底物组成的复合体的研究成果、X 射线研究成果（弄清楚了氢键形成位置）及其他工作。

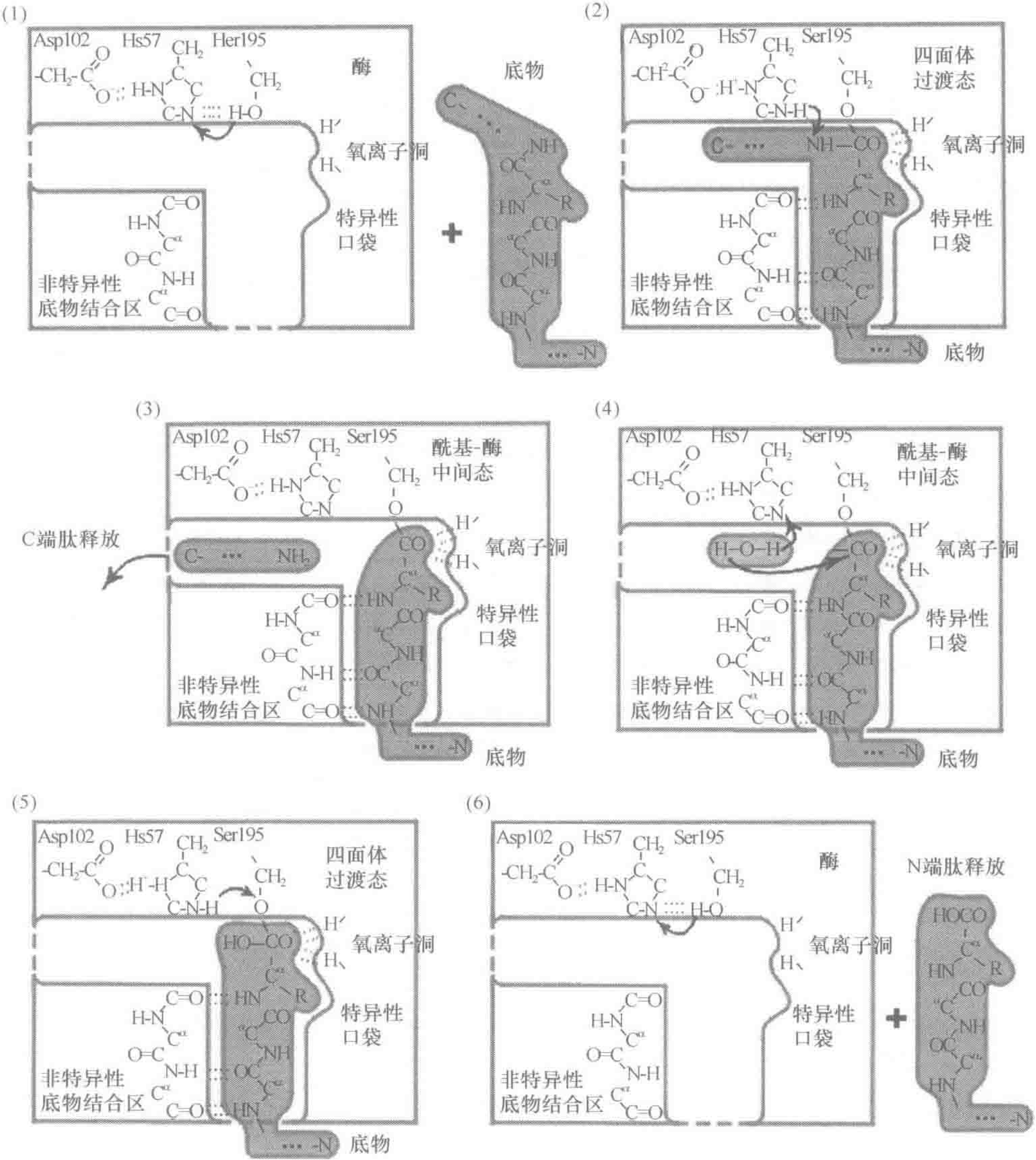


图 24-9 肽的酶促水解示意图

■ 那么，酶为加速化学反应究竟做了些什么呢？要回答这个问题，让我们来对比一下酶促反应（图 24-9）和另一个的类似的在水中进行但没有酶帮助的自发反应（图 24-10）。

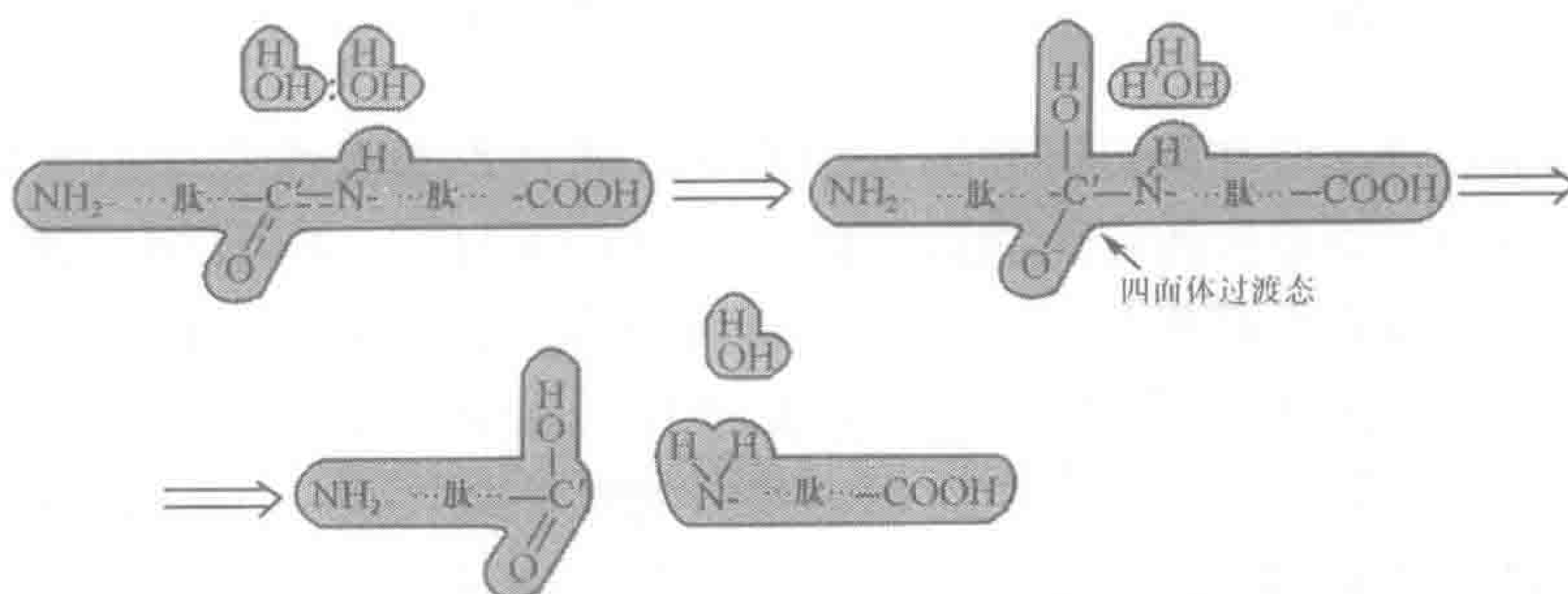


图 24-10　肽在水中进行的无酶助自发水解反应示意图。化学机制与图 24-9 类似

从图中可见，酶促反应是分两步进行的：首先是 C 端肽链断开，N 端肽链与丝氨酸的 O 原子形成一个复合体，然后这个 N 端肽把它与丝氨酸 O 原子的结合改换为与水分子的 O 原子进行结合。无酶助反应是一步进行的：C 端肽链断开，N 端肽链与水分子的 O 原子结合为复合体。我们还看到，这三个反应都是经由一个活化四面体复合物进行的，确切地说，是经由其四面体 C 原子（与 O 原子形成了一个共价键）邻近某个质子供体（在酶中是 His^+，在水中是 OH_3^+）的活化复合物进行的。

酶促反应过程中有两个活化复合物（无酶助反应中只有一个），这不是加快反应速度的因素。反应加速的原因在于，与酶结合的活化复合物比水中的活化复合物更**不**稳定（正是它的不稳定性限制了反应速度）。

在酶中，活化复合物（或者说，“过渡态”）以如下方式得以稳定化。

首先，C′原子的四面体化在肽链中产生了 C′—O^- 基团，其负电荷被吸引到酶中含有两个质子的氧离子洞中。这些带正电的质子与带负电的 O^- 原子之间的氢键和静电作用降低了后者的能量，并随之降低了 C′原子四面体状态的能量（而在无酶反应中，没有这种事先形成的“洞”），也就是说，导致过渡态的能量（焓）降低，并发生了所谓的**焓催化反应**（enthalpic catalysis）。

其次，邻近四面体的 His^+，其 C 原子带有质子，它几乎与不带质子的 His^0 一样稳定。在水中，提供质子的是 OH_3^+，由于其不稳定性，该离子的浓度在水中非常小，大约 10^{-7}mol/L。把它从水中“捞起”会降低熵，并相应地提高无酶活化复合物的自由能。因此，熵阻碍了活化复合物的各个成分在水中“成群”聚集，但并不阻碍它们在酶中聚集，它们就在那里组装起来。“捞起”底物的熵被它们黏附到底物结合口袋所弥补。这就易化了活化复合物各种成分的聚集，降低了它们在酶中的自由能（与无酶反应相比较），导致所谓的**熵催化反应**（entropic catalysis）。

总之，催化反应的焓成分和熵成分使酶促反应加速到无酶反应的 10^{10}倍。

如果底物与酶的黏附相当强，而底物的浓度又不太低，则熵催化反应可降低所观察到的反应对底物浓度的级数比。例如，如果无酶合成反应 $A+A \longrightarrow A_2$ 的速率正比于底物 A 浓度的平方，则类似的酶反应（反应第一步是底物 A 与酶相结合）依赖于酶的浓度，而不依赖于溶液中底物（或多种底物）的浓度。这种效应在反应涉

及几种分子和底物时(如在无水情况下由几个片段合成多肽)特别显著。

■　在化学催化反应和酶促催化反应中的基本特征都是过渡态自由能的降低,即在反应过程中需要克服的自由能最大值降低(图 24-11)。这个自由能的降低可以是由于反应的所有必需成分黏附在酶上造成熵损失而达成。而且,自由能的降低也可以是参与反应的分子的过渡态[而非其初始态(底物)和/或终末态("产物")]的优先结合所消耗的自由能而达成或增强。顺便说一下,如果底物或产物与酶结合过于紧密(超过过渡态),酶会被它们完全抑制,中毒(它会失去其"释放"功能而退出工作)。

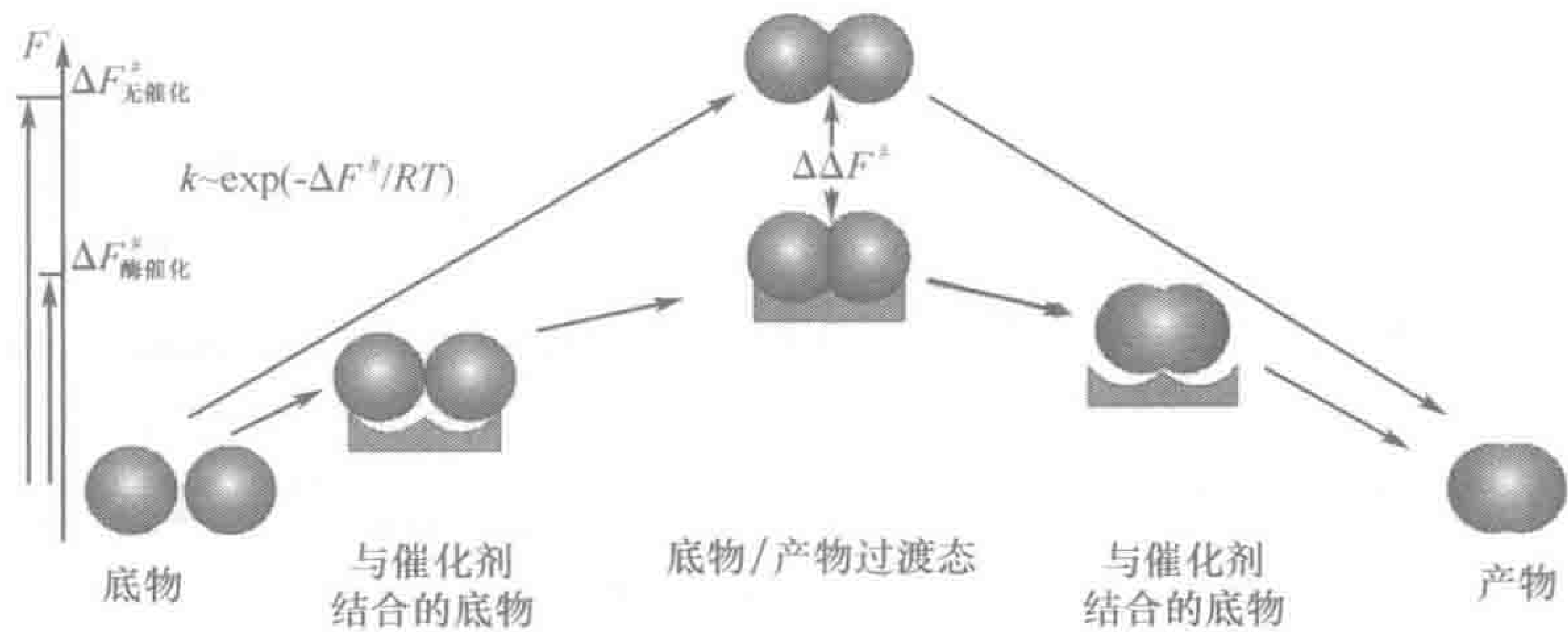


图 24-11　阐释底物/产物的过渡态(反应过程中最不稳定的状态)中催化剂稳定化作用的重要性和催化剂刚性对于降低反应自由能垒($\Delta F^{\#}$)、加快反应速度 k 的重要性的示意图。为了直观起见,图片只显示了过渡态外形与酶活性中心外形的互补性,在实际情况下是底物/产物的过渡态与酶催化中心之间的临时共价键发挥主要作用。催化剂并不改变初始态和终末态的自由能(F)。图中强调的是,反应的加速取决于 $\Delta\Delta F^{\#}$ 的大小,即取决于酶与过渡态结合的紧密程度

在过渡态的优先结合中,主导作用是发生了形变的电子系统(回忆一下丝氨酸蛋白酶中的"氧离子洞")中的结合,或者是所有分子的形变构象(同样也是在过渡态)的优先结合。

■　后一种情况可以拿人造"抗体酶"(abzyme, antibody enzyme)的催化功能清楚地阐释。抗体酶是被选择用于与底物过渡态结合从而催化其化学转换的抗体(图 24-12)。有些抗体酶能够把底物结合中所获得的全部能量用于降低活化能垒,即把几乎所有能量用于所处理的共价键的形变。虽然现有的抗体酶不是很强效的酶(由于在抗体酶中还未能建立电子供体和受体,它们最大只能够把自发反应加速 5 个数量级,而不像自然酶那样可达 10~15 个数量级),它们能够被创造出来的可能性①证明了过渡态优先结合的重要性。

---

①早在 20 世纪 60 年代末就产生了抗体酶的相关理论,但抗体酶技术在 80 年代才取得了突破性进展,并投入实用。

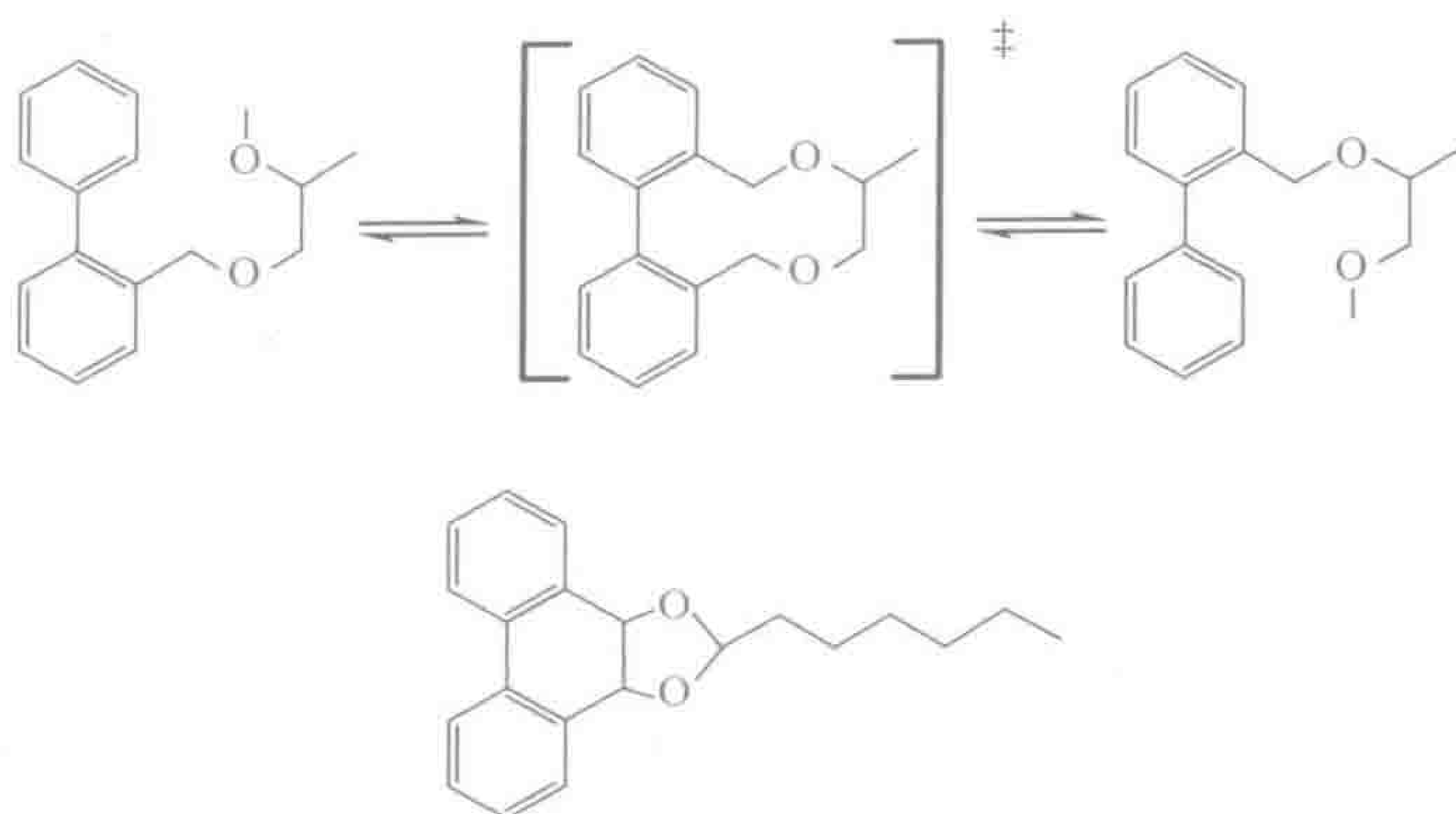
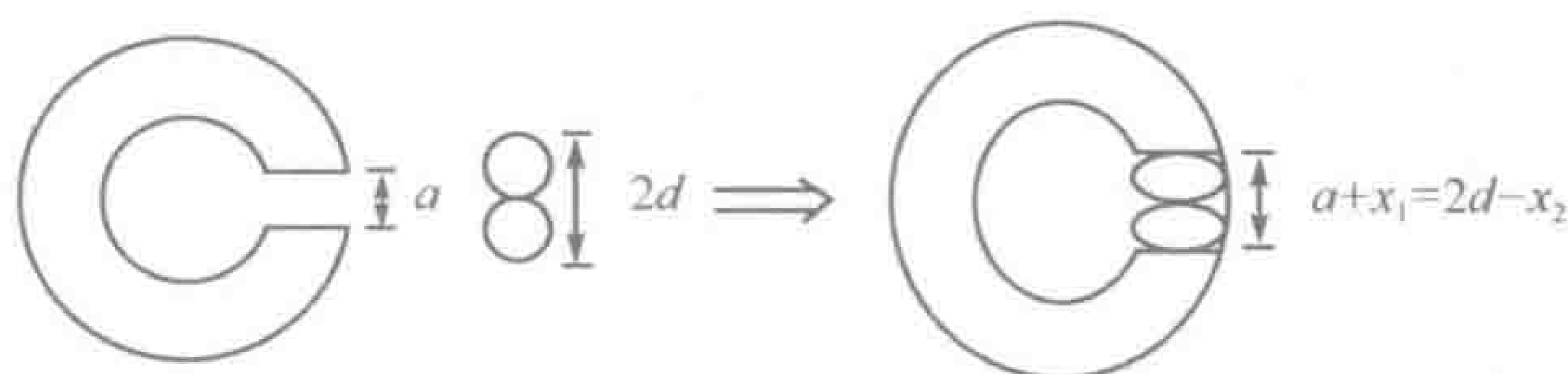


图 24-12　抗体酶(为缓慢自发反应的过渡态而人工选择的抗体)催化的化学转变原理图。图中给出的异构化反应的过渡态(‡)(与其初始态和终末态都不相同)是一个扁平的三环系统。为了制造针对这种形状的过渡态的抗体,使用了包含相似环、但稳定的分子(下图):这种分子被注射到动物血液中,在其中产生免疫性

我想要强调的是,为了恰恰与过渡态,与化学反应的起始状态和终末状态有几埃的峰值差别(图 24-11),进行优先结合,蛋白质应该尽量地坚硬。柔软的蛋白质,更确切些:柔软的活性中心,在催化过程中的作用就像用橡胶刀片刮脸一样……

也许,讨论下面这个简单模型是有益的。

假设有两个分子(下图中的两个小球,直径都为 d)以吸收能为代价被吸引进入到蛋白质的裂隙("活性中心")中,裂隙的初始宽度为 a(小于 $2d$)。在此过程中小球被压缩,而裂隙被扩大。蛋白质和小球被认为是有弹性的(后者有一定的压缩限度,超过之后会在它们之间形成共价键)。设蛋白质的弹性系数为 k_{pr},而两个分子小柱的弹性系数为 k_{mol}。求小球变形放出多少能量(仅仅只有这一能量可以用于克服在两个分子自发粘合为一个的反应途径中的能垒,如图 24-11 所示)。



解答如下。在小球的压力下,裂隙的宽度从 a 增大到 $a+x_1$,而两个小球组成的小柱的长度由 $2d$ 减小到 $2d-x_2$,且 $2d-x_2=a+x_1$。因此,根据虎克定律,用于蛋白质变形的能量为 $E_{pr}=1/2k_{pr}x_1^2$,而用于分子变形的能量为 $E_{mol}=1/2k_{mol}x_2^2$。而且有 $k_{pr}x_1=k_{mol}x_2$,即分子对蛋白质的压力等于(根据牛顿定律!)蛋白质对分子的压力。即有

$$\frac{E_{mol}}{E_{pr}}=\frac{\frac{1}{2}k_{mol}x_2^2}{\frac{1}{2}k_{pr}x_1^2}=\frac{\frac{1}{k_{mol}}(k_{mol}x_2)^2}{\frac{1}{k_{pr}}(k_{pr}x_1)^2}=\frac{k_{pr}}{k_{mol}} \tag{24.1}$$

和(进行一些计算之后)

$$E_{mol} = \frac{1}{2}k_{mol}(2d - a)^2\left(\frac{k_{pr}}{k_{mol} + k_{pr}}\right) \tag{24.2}$$

于是,我们看到,在“柔软的”蛋白质($k_{pr} << k_{mol}$)中,主要能量都花在对催化反应无用的蛋白质形变上了,而只有“刚性的”蛋白质($k_{pr} >> k_{mol}$)才能把主要能量集中在被处理的分子的形变上。

得证。

课堂里的声音:是否意味着,“刚性的”应该指的不是蛋白质本身,而是它的活性中心?而且,在讲到“刚性”时,您指的仅仅是蛋白质的力学刚性,或者其他什么的?

讲课人:我指的正是活性中心的刚性;蛋白质其余部分的刚性的重要性(对催化反应来说)仅仅在于它在多大程度上保障活性中心的刚性。而且我指的不仅是力学上的,而是能量观点上的刚性:即蛋白质在底物变化时不会微调,因而底物的状态微小改变会引起它与蛋白质相互作用的能量的巨大改变。这种效应可由底物周边与蛋白质的力学相互作用而引起;对此图片非常直观,如图 24-11 和图 24-12 所示。但是这种效应也有可能与底物和蛋白质的价电子相互作用有关;电子云相互作用的能量在其形状的非常“微小尺度的”改变下就会急剧变化。

底物的结构重建是化学反应的本质。与在较软的范德华相互作用下的底物变形相比,底物与酶催化反应中心的刚性临时共价结合能更有效地实现底物的刚性共价键的形变。这就解释了,为什么只通过与底物的范德华相互作用对反应起催化作用的抗体酶的催化效率,比带有化学反应催化中心的天然酶要低。

课堂里的声音:是否意味着,蛋白质的柔性在催化作用中不起任何作用?

讲课人:这里首先应该精确地区分催化活动本身(蛋白质的柔性不能促进它)和底物穿越进入活性中心的阶段,对此蛋白质的柔性是必需的;对后面这种效应我们还会遇到。其次,应该想到,如果蛋白质的活性中心不能与底物过渡态完美地契合(谁都不会没有缺点),则一些柔性能允许这种非理想的活性中心采纳哪怕是临时的形状。然而,重要的是,活性中心必须恰恰与底物的过渡态相契合,而不是与初始态或终末态。

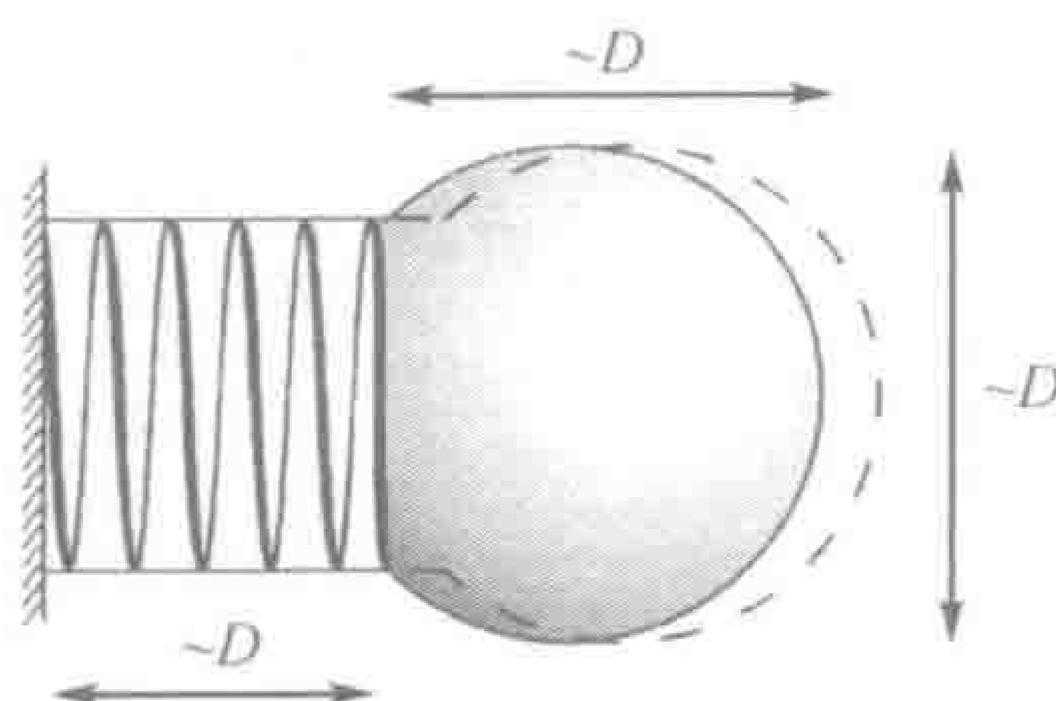
课堂里的声音:至今为止,您所有的推论只涉及静态的或者几乎静态的酶。对催化作用所需的能量能不能由酶的运动部分的动能(如它的振动)提供?

讲课人:不,不可能。可以证明,甚至是大蛋白质分子的动能都会在最多 ~10^{-11}s 内耗散掉(而小分子会更快),而一次酶促催化动作进行时间不短于 ~10^{-6}s。即,底物拼装到酶上所释放出的动能远在可以以哪怕任何一种方式被用在催化活动之前就耗散掉了,转化成热了。

实际上,在黏性的周围环境中(水、膜等)可以发生蛋白质的一部分相对于别一部分的运动(如在振动时)。这种运动可用平常的简单方程来描述:

$$m\frac{d^2x}{dt^2} = F_{frict} + F_{elast} \tag{24.3}$$

式中,d^2x/dt^2 为粒子运动速度;m 为其质量;F_{frict}是与周围媒质的摩擦力;F_{elast}是作用于其上的弹性力。质量可用 $m=\rho V$ 估算,其中 ρ 为粒子的密度,V 为体积。F_{frict} 可根据斯托克斯定律估算:$F_{frict}=-3\pi D\eta(dx/dt)\approx -10D\eta(dx/dt)$,其中 dx/dt 为粒子的速度,D 为其直径,η 为媒质的黏度。F_{elast} 可根据虎克定律估算:$F_{elast}=-(ES/L)x$,其中 x 为偏离平衡点的位移,S 为蛋白质两部分结合处的截面积,L 为粒子的弹性长度,E 为其弹性模量。在后续估算过程中可以合理地认为,所有线度大致是相同的(如附图所示),即有 $L\approx D, S\approx D^2, V\approx D^3$。



式(24.3)给出了两个特征时间。

一个是我们已经讨论过的最重要的特征时间 t_{kinet},是在与周围媒质摩擦作用下的蛋白质运动阻尼时间。这个时间的估算需要根据这个方程的"运动学"部分推算:$m(d^2x/dt^2)\approx -3\pi D\eta(dx/dt)\approx -10D\eta(dx/dt)$。得到的是 $t_{kinet}\approx m/(10D\eta)$;它很容易定量地进行估算。由于 $m\approx\rho D^3$,则 $t_{kinet}\approx 0.1\rho D^2/\eta$。而且,由于蛋白质的密度 $\rho\approx 1g/cm^3$,而水的黏性 $\eta\approx 0.01[g/(cm\cdot s)]$,则 $t_{kinet}\approx 10^{-13}s(D/nm)^2$,其中$(D/nm)$是以纳米为单位的蛋白质直径。这意味着,尺寸 $D=1nm$ 的蛋白质一小部分的动能在水环境中在 $10^{-13}s$ 内耗散掉,而对尺寸 $D=10nm$ 的大蛋白质,动能在 $10^{-11}s$ 内耗散掉。在黏性比水更大的环境(膜或蛋白质其他部分)的摩擦下,内能耗散得更快。

有必要再估算一下蛋白质振动的特征时间 t_{vibr}。它的数值可以从式(24.3)的"振动"部分推算:$m(d^2x/dt^2)\approx -Edx/t_{vibr}\approx(m/ED)^{1/2}\approx D(\rho/E)^{1/2}$。蛋白质的典型弹性模量为 $E\sim 10^{10}g/(cm\cdot s^2)$,$t_{vibr}\approx 10^{-12}s(D/nm)$。对比 t_{vibr}和 t_{kinet},我们看到,在尺度 $D<10nm$ 的蛋白质中,由于摩擦的缘故,弹性振动是完全不可能的(很快就阻尼了)。

这意味着:①势能只可能有效地储存在底物中(以便把它引进到过渡态),而不是在蛋白质中,因此蛋白质的活性中心必须是刚性的;②在催化反应中无法利用蛋白质的内能。

■ 以过渡态优先结合为代价的催化理论是早在20世纪30至40年代由Holden和Pauling提出来的。最近,通过蛋白质工程,即通过酶活性中心的氨基酸定向替换,多个酶促反应的过渡态被"探测"到了。而且成功地说明了(并通过使用X射线看到了)酶的哪些氨基酸负责黏合对形成活化复合物所必需的各个成分,哪些氨基酸负责恰好而且专一地与过渡态(以前它只是理论性的,因为那时还没在结构实验中观察到复合物的结构)优先结合。

例如,在Alan Fersh所研究的酪氨酰tRNA合成中,找到仅过渡态中与ATP γ

磷酸结合的氨基酸残基(Thr 40 和 His 45,如图 24-13 所示),这种结合就使该过渡态与简单结合的底物或反应产物区分开来。

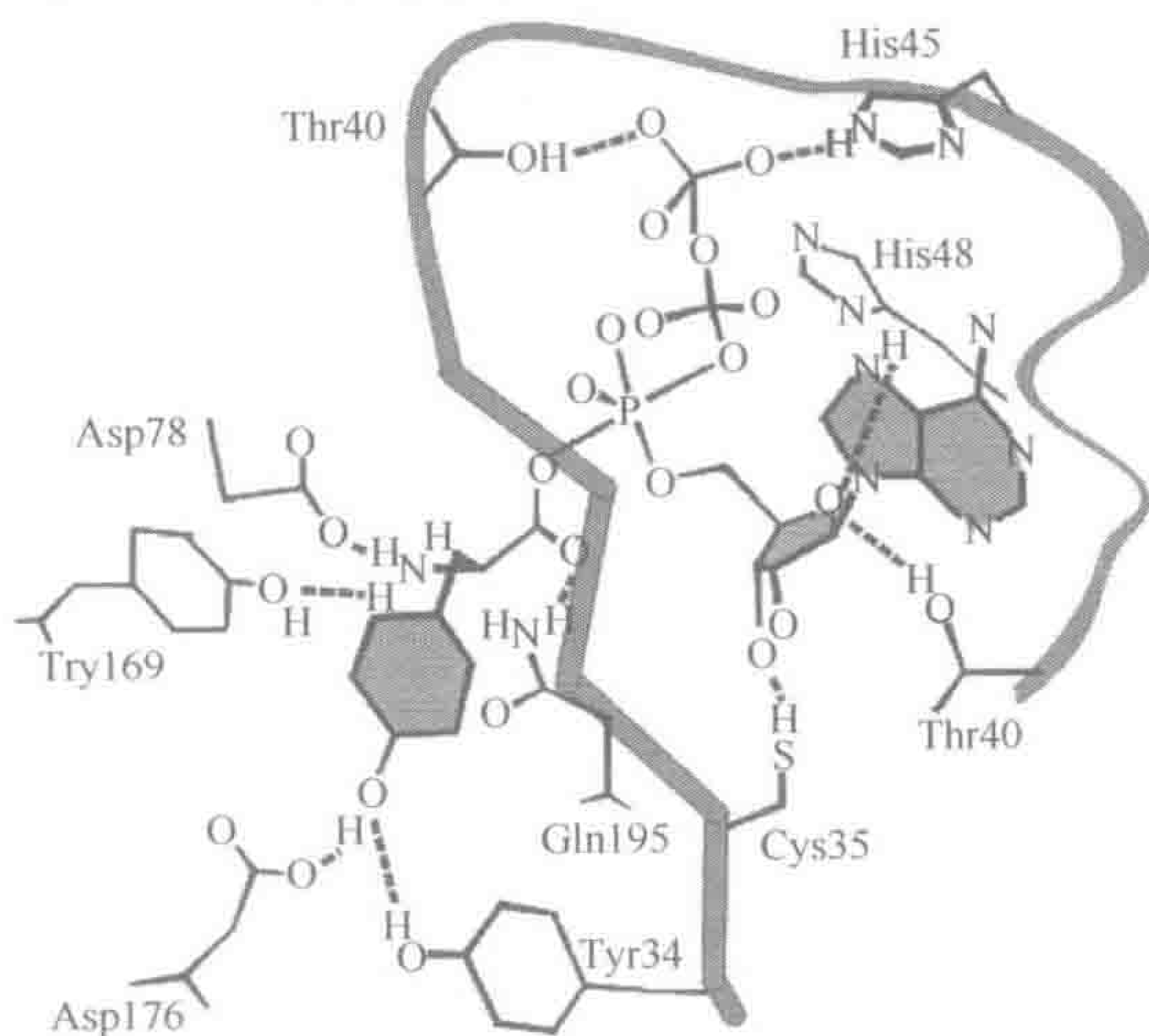


图 24-13 在酪氨酰 tRNA 合成酶上进行的由酪氨酸和 ATP 合成酪氨酰 AMP 反应中的过渡态的推测结构。为直观起见,酪氨酸的环画上了阴影,腺苷环也是如此。ATP 分子中攻击酪氨酸的活化五价 α 磷(P)位于图中心,而 ATP γ 磷酸在它的上方部分。这个磷酸只在反应的过渡态与 Thr 40 和 His 45 形成氢键;在非产物与底物结合的简单情形下,没有这些氢键。图片取自文献[9],略有修改

过渡态极不稳定(根据定义:最不稳定的!),因此它不可能被直接观察到。X 射线只能观察到底物在酶上的稳定分布,看不到它(底物)与 Thr 40 和 His 50 的结合(这时距离太大了)。把 Thr 40 和 His45 替换为小的 Ala 能把突变后的蛋白质的活性降低到百万分之一以下,尽管这些替换几乎不会而影响"突变体"对反应底物的结合。这样就证实了它们(Thr 40 和 His 45)与 ATP γ 磷酸的结合只发生在过渡态:攻击酪氨酸 CO 基团的 α 磷的应变态(由于它具有五个共价键)。

■ 讨论酶促反应时,我们不能不注意到它的高度特异性。

例如,丝氨酸蛋白酶只在特定的氨基酸之后切割多肽:胰凝乳蛋白酶只在芳香族氨基酸之后,胰蛋白酶只在带正电的氨基酸之后,弹性蛋白酶只在最小的氨基酸之后。这种效应[它被称为根据"钥匙(底物)-锁(酶)"原理的底物识别]是通过底物结合"口袋"(图 24-14)的**特异性**部分的巧妙结构来实现的,在该处应该放得下剪切点之前的最后一个(如在胰蛋白酶中)或倒数第二个(如在木瓜蛋白酶中)侧链基团。在嗜热菌蛋白酶中特异性的钥匙是位于紧接着被剪切的 C-N 键**之后**的氨基酸残基。

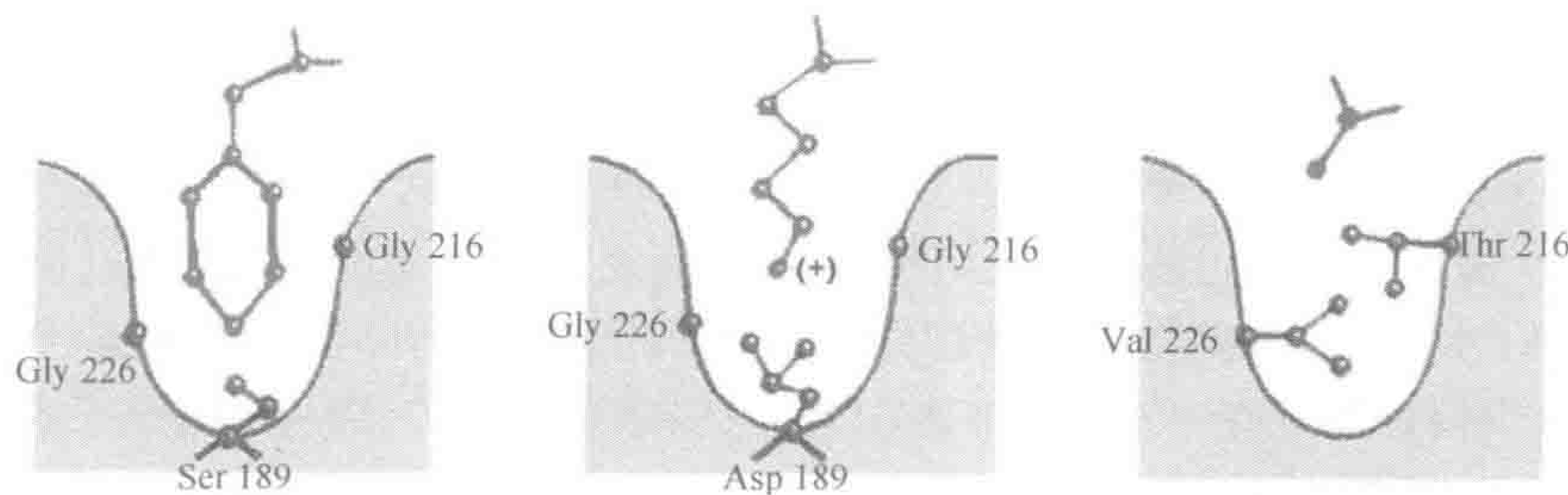


图 24-14 各种丝氨酸蛋白酶的底物结合口袋中的特异性位点的结构。图片经许可取自文献[5]

关键残基相当准确地标出了可能的切割点，但多肽在标定点的切割速率还依赖于该点周围的其他残基（这是很自然的：因为它们也多少与蛋白质发生一些相互作用）。

然而对于切割效率和切割可能性本身还有更重要的因素，那就是多肽在切割部位的构象。如果该构象不利于切割（即如果关键侧链基团和有待切割的 C-N 键不能在酶上同时占据“正确”位置），则这个 C-N 键不会被切断。

因此蛋白酶在切割具有刚性空间结构的蛋白质时比较差，但是，只要当这些蛋白质发生变性，就很容易把它们摆平。

如果多肽链具有某些特定的刚性构象，甚至可以只是黏附在蛋白酶上，封阻它的活性中心，一旦如此，则不会被切断，而总是占据在它上面。许多抑制剂就是以这种“无效结合（unproductive binding）”原理工作的。

今天我们讨论到的蛋白酶是很有效的，但并不是绝对精密的。看来，生物体还可以允许它在选择切割点时偶尔犯点错。但是有些蛋白质要求绝对可靠。这种可靠性是通过在同一个蛋白质上的几个活性中心的工作来达到的。

今天讨论了蛋白质的基本功能之后，我们将在下一讲中考虑它们更为复杂的功能。

第 25 讲

功能的组合。底物从一个活性中心到另一个活性中心的转移。“双层筛”增加功能特异性。蛋白质结构对它的基本酶活性的相对独立性。蛋白质结构与其环境的可见联系。蛋白质基本功能与其结构的匹配。诱导契合。蛋白质域的活动性。蛋白质域在进化中的移动。域结构:激酶,脱氢酶。变构:阳离子中心的相互作用。蛋白质功能的变构调节。变构与蛋白质四级结构。血红蛋白与肌红蛋白。肌肉收缩机制。肌丝行走。

上一讲中讨论了蛋白质的基本功能,我们现在来考虑它们更为复杂的功能。

■ 如前所述,有一些蛋白质需要有极大的工作精确度,而这样的精确性是通过同一蛋白质中的几个活性中心的耦合工作来实现的。例如,在氨酰 tRNA 合成中就是这样的情形:它们用合适的氨基酸使 tRNA“充电”,而且氨酰 tRNA 合成的精确度依赖于这些氨基酸工作的精确度。因此它们不能容许在数千个氨酰 tRNA 合成中出错超过一次,而氨基酸对特异性底物结构口袋的识别允许出错率约为 1%。为了消除这些错误,使用了“双层筛”效应(图 25-1)。这种效应是由酶的特殊结构实现的。

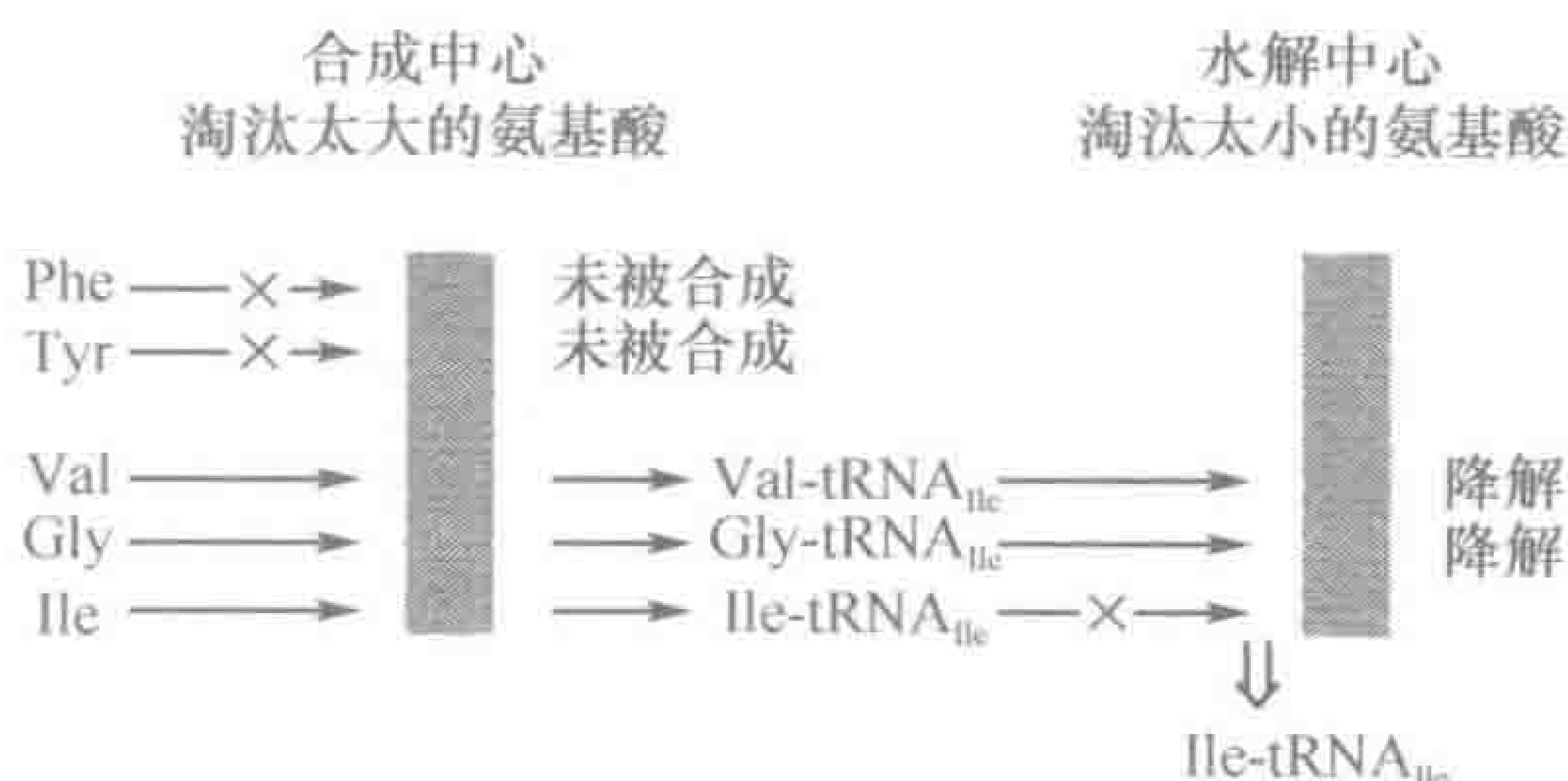


图 25-1 异亮氨酰 tRNA 合成工作(用异亮氨酸(Ile)使异亮氨酰 rRNA($tRNA_{Ile}$)带电)中的校准机制“双层筛”示意图

氨酰 tRNA 合成酶有**两个**活性中心:合成中心和水解中心。此外,氨酰 tRNA 水解**不是**一个与氨酰 tRNA 合成刚好相反的反应:这两个反应进行时都释放游离磷酸根,即**两者**都伴随有自由能降低。

而且,氨酰 tRNA 水解中心的底物结合口袋比合成中心的口袋要小。

筛子的原则是:**不**允许比筛子网孔大的粒子通过。

在氨酰 tRNA 合成过程中的第一层筛子处,tRNA 被“正确”的氨基酸和一些较小的“错误”氨基酸所充电(根据疏水性/亲水性有差别的原理,更大的氨基酸全部被拒绝,而大部分较小的“错误”氨基酸也被拒绝,但完全根据疏水性/亲水性进行

的挑选没有根据能否进入特定尺寸的口袋而进行的挑选那么严格)。因此,在合成阶段的输出中,大部分氨酰 tRNA 带有“正确”的氨基酸,少量氨酰 tRNA 带有“错误”的氨基酸,但只有比“正确”氨基酸要小的:更大的氨基酸被合成中心的“口袋”(第一层筛子)所阻挡。

在第二个活性中心,即第二层“筛子”上的“网孔”(底物结合口袋)比较小。在该处发生已形成的氨酰 tRNA 的水解反应,但**只**水解那些带有比“正确”氨基酸要**小**的氨基酸的氨酰 tRNA。因此“正确”氨基酸被水解中心的口袋,即“第二层筛子”,所拒绝,而所有其他氨酰 tRNA 都被水解并分裂掉。

总之,这两层筛子保证只输出“正确带电的”氨酰 tRNA。

■ 在结束介绍酶的特异性时,值得介绍一下两个实用主义观点。第一,目前有很多人在研究底物结合口袋并挑选,有目的地,而非盲目地,与它们相结合的抑制剂。出于可以理解的原因,其中特别感兴趣的是艾滋病病毒(HIV)的蛋白酶和其他蛋白质。第二,活性中心(特别是底物结合中心)的点突变得到了大量应用:它们能够改造“天然”底物的特异性(如改变丝氨酸蛋白酶的特异性),甚至创造新的对于农业和医药业所需的人工特异性。也有很大的兴趣在尝试(到目前为止仅仅是尝试)通过对原始蛋白质的氨基酸残基进行定向点替换,创造(在蛋白质球体的“坚硬底座”上)具有 “嫁接的”新功能的“新”活性中心。

■ 对蛋白质的酶活性的研究表明(请回忆上一讲的内容),蛋白质球体中只有一小部分积极参与到这种功能中,而其余部分只是作为坚实的骨架,以保证固定在其中的活性中心的“正确”结构。

因此,我们不应该感到奇怪的是,具有完全不同的一级结构、甚至具有完全不同的空间结构的蛋白质,可以具有相同的或非常相似的生化功能。

经典的例子又是丝氨酸蛋白酶。这种蛋白酶有两类:胰蛋白酶型和枯草杆菌蛋白酶型。它们既在氨基酸序列方面,也在总体形状上完全**不**相似(图 25-2),甚至属于不同的结构分类(胰蛋白酶是两域的 β 型蛋白质,枯草杆菌蛋白酶是单域的 α/β 型蛋白质)。它们只是催化中心的关键氨基酸残基有相同的构象(图 25-3),而且并非是全部残基的整体构象,而只是它们的功能上重要的“末端”的构象相同,甚至在底物结合口袋里也没有相同的残基构象。

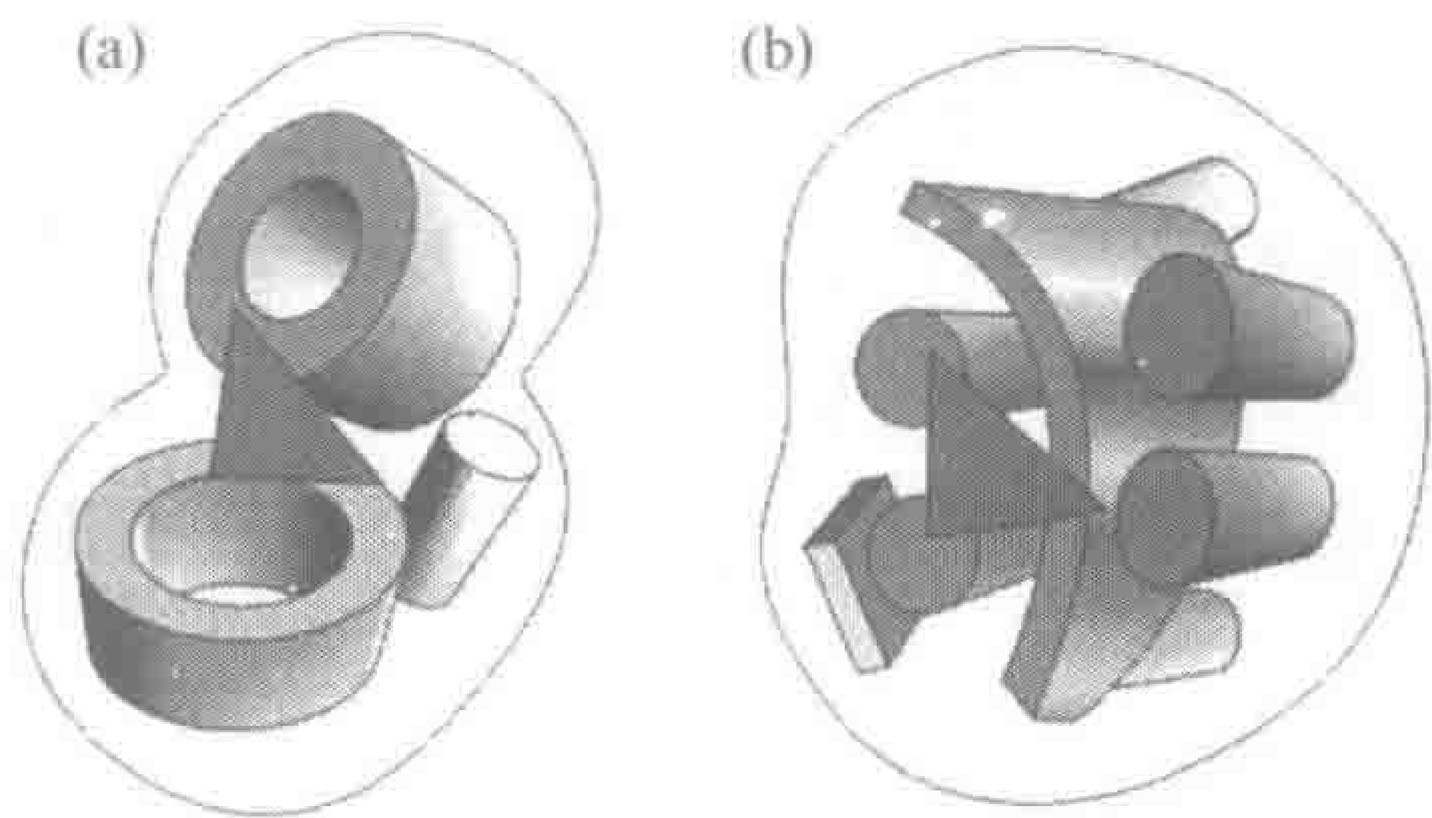


图 25-2 胰蛋白酶型(a)和枯草杆菌蛋白酶型(b)丝氨酸蛋白酶的结构示意图。在球体基本轮廓的背景上显示了 α 螺旋和 β 片(比较扁平或扭转成桶)。活性中心区域显示为黑色三角形

此外,还有一些蛋白酶,甚至在催化中心区域也完全没有丝氨酸蛋白酶中的相

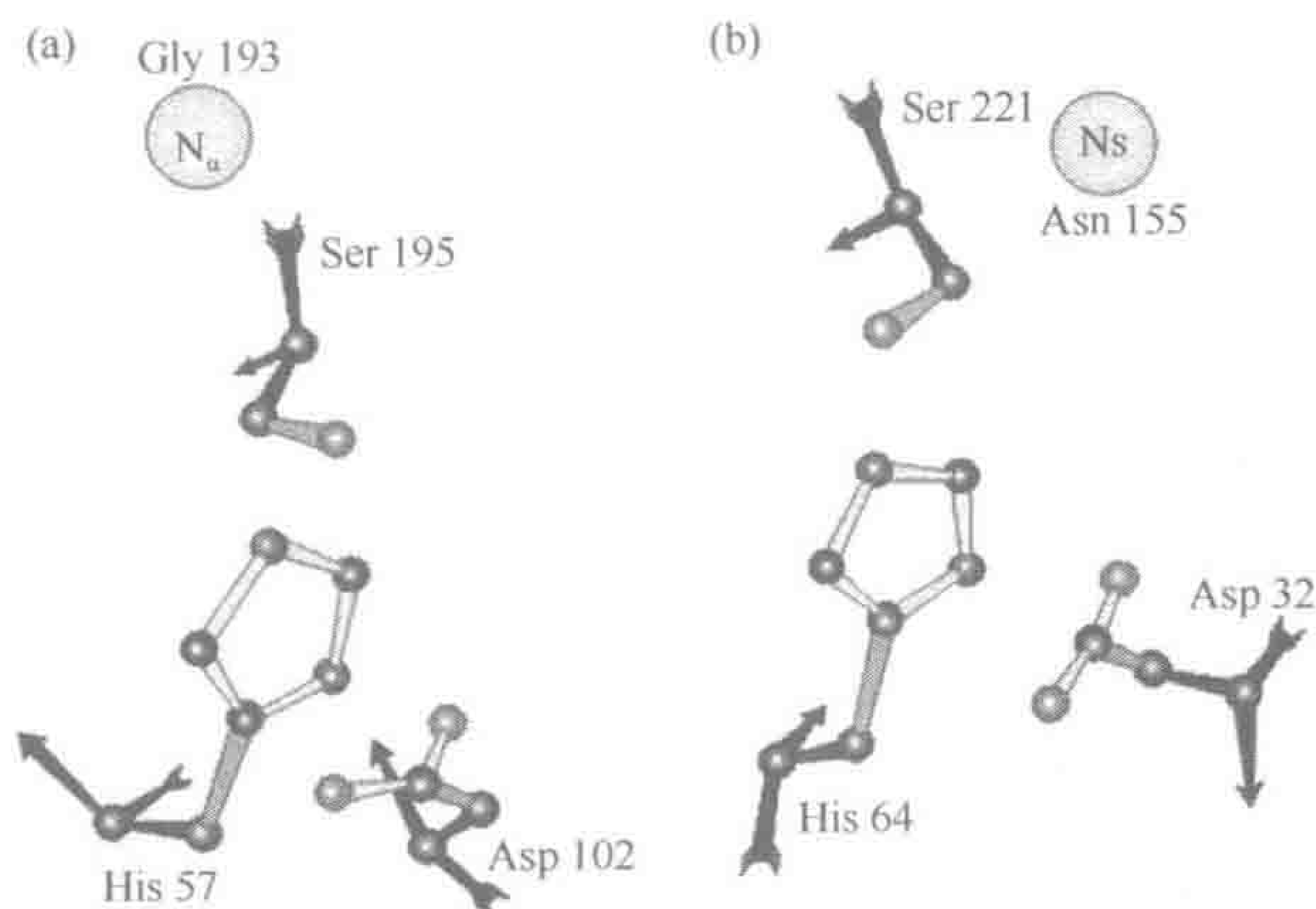


图 25-3　胰蛋白酶型(a)和枯草杆菌蛋白酶型(b)丝氨酸蛋白酶的催化中心的结构示意图(顺便说一句,相似的催化中心不仅存在于蛋白酶中:它们还存在于脂肪酶和其他一些酶中)。主链片段用黑色标出(直到 C^β 原子),这些片段中的链走向路径以箭头显示。圆圈(标有 N)表示氧离子洞的位置。在胰蛋白酶中主链的 N_α—N 基团面对着它,在枯草杆菌蛋白酶中侧链的 N_δ—H 基团面对着氧离子洞。可见,在这两类丝氨酸蛋白酶中,只有转运电荷的催化三联体(Ser,His,Asp)的侧基团末端的分布是不变的,氧离子洞有点勉强地说也是不变的。而甚至在关键催化残基 His 和 Asp 的主链走向路径中也能看出有很大差异

似性。我说的其中包括羧肽酶型和嗜热菌蛋白酶型的金属蛋白酶。在催化作用中它们中起关键作用是紧密嵌入到蛋白质中(并在该处保持配体键)的 Zn^{2+}。带多电荷的金属离子由于失去外部电子而具有较小的半径,它们在自身周围建立很强的电场。这使得嵌入在金属蛋白酶中的离子能使水分子活化(把它分解成 OH^- 和 H^+),并进而使肽键断裂。在这种过程中活化水大致像我们在上一讲中已经讲过的非催化反应中一样“工作”。但本质上的差别在于,酶恰恰只活化那些处于被攻击的肽键附近的水分子,而这在非催化反应中需要在大量非活化水分子中长时间寻找才能发生。同时,在金属蛋白酶中,Zn^{2+} 的电荷能使所分裂的多肽的四面体过渡态上的负电荷稳定化(即模仿丝氨酸蛋白酶中的氧离子洞的作用)。

■　因此,同样的化学反应可以由完全不同的蛋白质来完成。而且,起到同样的关键作用的(强力的电动切割机的作用)在一些蛋白质中是带多电荷的金属离子,在另一些蛋白质中是容易接受电子或质子的有机辅因子,在第三类蛋白质中是由自身周围所活化的侧链基团……

另一方面,正如我们记得的,具有 α/β 圆筒特征形状(或具有 Rossmann 折叠形状)的不同蛋白质可以催化完全不同的反应。

一些蛋白质具有共同的结构和共同的起源,但是功能不同。经典的例子是溶菌酶与 α 乳清蛋白。这两个蛋白质不仅具有几乎相同(精确到 1～2Å)的空间结

构,而且是同源的,它们有 35%完全相同的氨基酸残基。不过,其中一个(溶菌酶)是酶(切割寡糖),而另一个(α 乳清蛋白)是不久前(1 亿年前,牛奶出现的时候)从溶菌酶起源的,它失去了溶菌酶活性中心,现在本身完全不再是酶,而是修饰另一种蛋白质的酶活性。

■ 对蛋白质的酶学性质[按酶学分类“EC”(enzyme classification)]与其结构分类(CATH)之间的关联分析表明,总体上,不存有蛋白质的某种结构类型对某种生化功能的明显偏好。

但是,蛋白质“更大尺度的”性质(不是依赖于一个小小的催化中心,而是依赖于更大的球体表面),如结合各种大配体的能力,是与蛋白质的总体架构部分相关的。而且蛋白质对各种生物学过程的参与程度也部分地与蛋白质的总体架构有关。

最明显的是体现在负责跨细胞膜运输(运输对象种类极多)的蛋白质。作为膜蛋白,它们都有非常规则的跨膜 α 螺旋或 β 结构;对此我们已经讲过。我们也讨论过纤维蛋白,它们由于有巨大的二级结构模块,主要是起结构性的作用。不过,这种关系也体现在水溶性球状蛋白中,只是在较小程度上。

例如,血红素结合蛋白和 DNA 结合蛋白有半数是 α 型蛋白,而在糖结合蛋白和核苷酸结合蛋白中 α 型蛋白的比例不超过 20%。另一方面,几乎所有负责免疫和细胞识别的蛋白质都是 β 型蛋白。将近一半的糖结合蛋白也是 β 型蛋白,而在血红素结合蛋白、核苷酸结合蛋白和 DNA 结合蛋白中 β 型蛋白的比例仅约为 10%。几乎所有的核苷酸结合蛋白都属于 α/β 型蛋白。其中包括如糖酵解途径中的全部 11 个蛋白(它们使用核苷酸作为辅因子和底物)。一般来说,酶,尤其是那些维持细胞“内部家务”的蛋白质,常常是 α/β 型蛋白(多于其他类型)。

所有这一切都表明,虽然蛋白质的架构与其“大尺度的”底物结合活性、周围环境(“生活条件”)和对生物过程的参与度有一定关联,但是蛋白质的总体结构与其酶活性是相当宽泛地不相关的。

■ 到现在为止,我几乎没有说到蛋白质分子在行使其功能的过程中的构象变化,这不是偶然的:我们至今为止讨论的是孤立的酶反应,蛋白质构象的柔性只可能使其催化性质变坏。实际上,对有效的催化作用所必需的恰恰是对过渡态(它与初始态和终末态最大差别也只有 1Å)的优先结合,而且这种优先结合的目的在于合成或断裂非常坚固的共价键。而用“柔软的”蛋白质去撕裂它们——就像用橡皮或枕头去切割电线一样……

不过,这里我们必须作如下补充。当蛋白质的活性中心没有完全调整到适应它们的功能时,而理想化的调整需要自然选择的超精细工作,该中心可能需要一些变形以转变到活性形状。而如果这种变形确实是功能性必需的,它将会时不时地

引导不完善的蛋白质达到要求。不过这里应该明确区分责任(补偿活性中心几何缺陷的变形)与美德(蛋白质的完美性)。

当然,即使是最好的蛋白质在其催化工作过程中也会稍许变形,因为它不是钻石,但是这种作用与纸张在钢笔尖下的变形大致相同:越小越好。

课堂里的声音:那么,在酶催化中心完全不应该有柔性的细节?

讲课人:为什么完全不应该呢?只有当酶需要快速工作的时候,它们不应该有。而在G蛋白的催化中心,有这样的柔性细节。但是,G蛋白是干什么的?它的作用是分解GTP,但是——**非常缓慢**!实际上G蛋白(我们已经讲到过)是定时器。对于**慢速**的工作,柔性的催化中心非常适合……而对于"快速"的催化需要刚性的活性中心。

■ 当蛋白质需要从一种行为转换到另一种行为时,那就是完全另一回事了:为保证它的角色转换而作出的形变,实际上具有重要的功能性作用。

例如,在许多蛋白质中,底物进入到活性中心需要邻近该中心的蛋白质基团略微移开。但是当底物坐落在该中心上时,催化工作就不再需要这些基团的移动。

可以说,灵活性对为反应进行准备时所必需的,而刚性是在反应进行过程中所必需的。这就像军事上的"分进合击"原则。

不过,有时候蛋白质从一种功能到另一种功能的转换只是简单通过底物从一个活性中心转移到另一个活性中心而完成的(请回想氨酰tRNA合成酶:在那里所有移动都是底物作出的)。

然而,从一种基本功能转变到另一种基本功能常常是通过蛋白质结构或多或少大尺度的变换完成的,而且这种变换的必要性反映在蛋白质分子的结构上。对这一主题我们将在后面涉及。

■ 在讲述蛋白质的运动性时,必须区分小运动和大运动。

小运动,本质上就是热涨落。高分辨率的X射线结构分析能帮助我们在被平均后的"静态"蛋白质结构的背景上看到这种涨落。即根据原子位置的某种模糊度看到这种涨落,而且模糊程度正如热波动决定的那样,随温度增大。

蛋白质表面基团的波动较大(其振幅平均约0.5Å),而位于蛋白质核中的基团的波动只有几分之一那么大。环线上的氨基酸残基振动得比二级结构中的更强烈。侧基团振动比主链明显更强烈。较长的表面侧基团的振动特别强烈,由于它们的末端没有被其他基团所抓住,因而经常运动得太强烈以致X射线完全无法把它们记录下来。所以有时候人们说,蛋白质内部是固态的,而其表面像液体,只有到大约-100℃才凝固。(我顺便提及一下:半液体状的表面层是晶体的一般特征。)

γ共振谱("穆斯堡尔谱")也得出结论:在蛋白质球体的动力学中,"坚硬的"

和“松弛的”成分组合在一起。这一方法通过用某些重金属同位素(如 Fe^{57})的“穆斯堡尔”核监测单色 γ 量子的吸收,可以估算,尤其是,特别强烈地与这些穆斯堡尔核胶接的蛋白质固态部分的尺寸。

更大尺度的改变与蛋白质结构在与配体结合时的变形有关。这种变形的弛豫动力学已在肌红蛋白中得到了很好的研究。激光闪光几乎在瞬间($\sim 10^{-13}$s 内)把配体 CO 从其血红素上撕下来,然后监视(根据光谱)直到恢复到初始状态。

在宽广的温度范围内对这种弛豫的复杂动力学进行的研究表明,通向天然构象的路径要穿过一系列能垒,而蛋白质天然构象本身包括一套在微观细节有差别的亚构象。

蛋白质在行使其功能时的这些以及更大尺度的形变还可通过把蛋白质在不同功能状态下结晶进行研究。各种光谱法、化学修饰等等也提供了丰富的信息。

■ 我们已经在上一讲中已经了解到以色氨酸阻遏蛋白调控 DNA 结合活性为例的一种功能上重要的形变。现在我们来更细致地考虑功能上重要的形变的问题。

首先,让我们来看蛋白质的形变是如何帮助把“**结合→变换→释放**”这一循环的各个阶段联合起来的。图 25-4 所示为磷酸化蛋白质,己糖激酶,对其底物的**诱导契合**(induced fit)(D. Kosh-land 提出的术语)的示意图。这种蛋白质从 ATP 上转移磷酸基团到葡萄糖。原则上,根据化学反应,这种磷酸基团也可以转移到水中;但实际上并没有这样发生。在努力解答为什么没有这样发生的原因时,Koshland 提出以下假设:①在与底物结合之前,酶处于“开放”形态(这时它可以从水中抓取底物,但不能进行磷酸化反应;②在与底物结合之后,酶转变为“封闭的”催化活性形态,活性中心的所有部分组合成一个整体并能催化磷酸化反应,但水被排出了活性中心,并因此不能与底物竞争磷酸化;③在催化活动之后酶重新打开,放出已被磷酸化的底物。

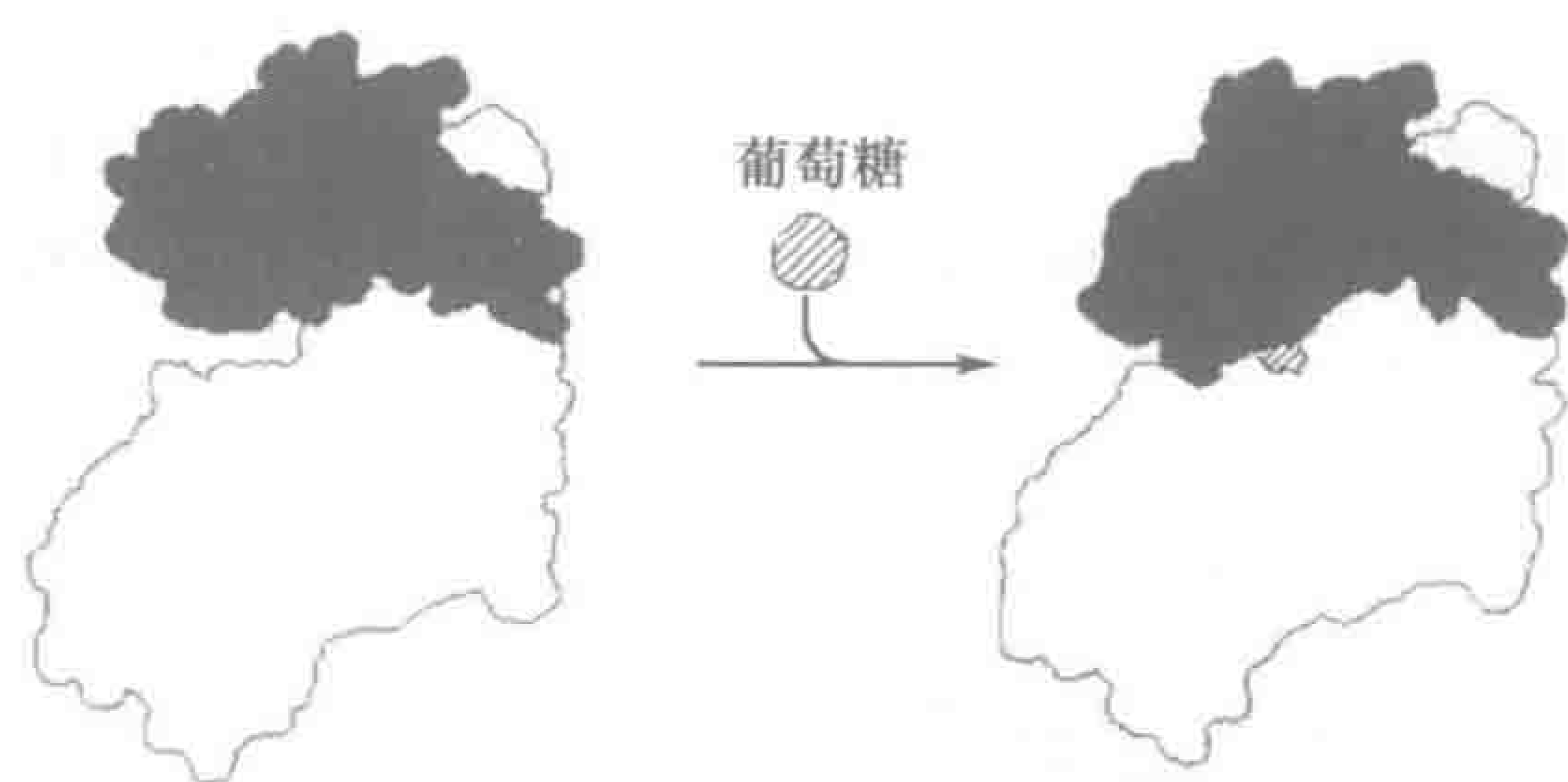


图 25-4 己糖激酶磷酸化过程中的**诱导契合**。在开放形态中,两个域被较深的裂缝分开,葡萄糖可以进入其中。当葡萄糖进入裂缝之后,域发生转动,裂缝被关闭,水从中排出,催化中心的所有部件组合到一起。图片来自文献[5],略有改动

随后的实验充分地证实了这一假设(图 25-4),但**只是**对于那些需要把所处理的底物和与之竞争的水隔离开的蛋白质(虽然有过许多试图把诱导契合机制扩展到所有酶反应的尝试)。例如,对胰蛋白酶的活动,这就不需要,而且也没有观察到在它里面对底物的诱导契合:胰蛋白酶(以及胰凝乳蛋白酶、弹性蛋白酶、枯草杆菌蛋白酶等)不发生形变,是根据简单的“钥匙-锁”原理识别底物的。

我想请大家注意,诱导契合是通过蛋白质大模块(请回忆上一讲)或整个蛋白

质域的**移动**、而**不**是整个蛋白质链折叠形状的重排(rearrangement)来达成的(图25-4)。而这些移动主要是通过微小的局部变形来进行的。[类比:肌肉收缩("局部变形"),让手指("域")压缩进入拳头,但手指并没有变成牙齿或触角……]

这也同样适用所有其他已知的"蛋白质构象重排"情形,除少数情况下,当整个 α 螺旋或 β 串从球体中伸出并形成不规则的构象时例外。我知道的最大重排之一发生在钙调蛋白(calmodulin)中 。它本身具有哑铃形状,α 螺旋域的"头"被"手柄",一个长 α 螺旋,分开很大距离。然而在它与其他一些蛋白质结合时,原先的哑铃"头"没有被打破,而是相互靠近并和蛋白质靶一起粘合起来,而原先的"把手"(长 α 螺旋)被拆断。

■ 蛋白质的域不仅在空间中是可移动的,而且在进化时间中发生移动。正如我讲到过的,域的基因作为一个整体,可以从一种蛋白质漫游到另一种蛋白质,有时候以不同的组合时而与其他域相结合,时而分离。常常发生这样的情形:在一种生物体中观察到的几种单域蛋白质,在另一种生物体中组合成为一个多域蛋白质。

域的相对自主性在被称为脱氢酶的蛋白质大家族中很容易看出来。这类蛋白质借助于 NAD^{+}/NADH 辅因子(它们很容易接受和放出单独的质子)催化类似于 OH 基团氧化的反应。不过它们处理的物质各不相同:乙醇脱氢酶—乙醇(酒精),乳酸脱氢酶—乳酸,依此类推。

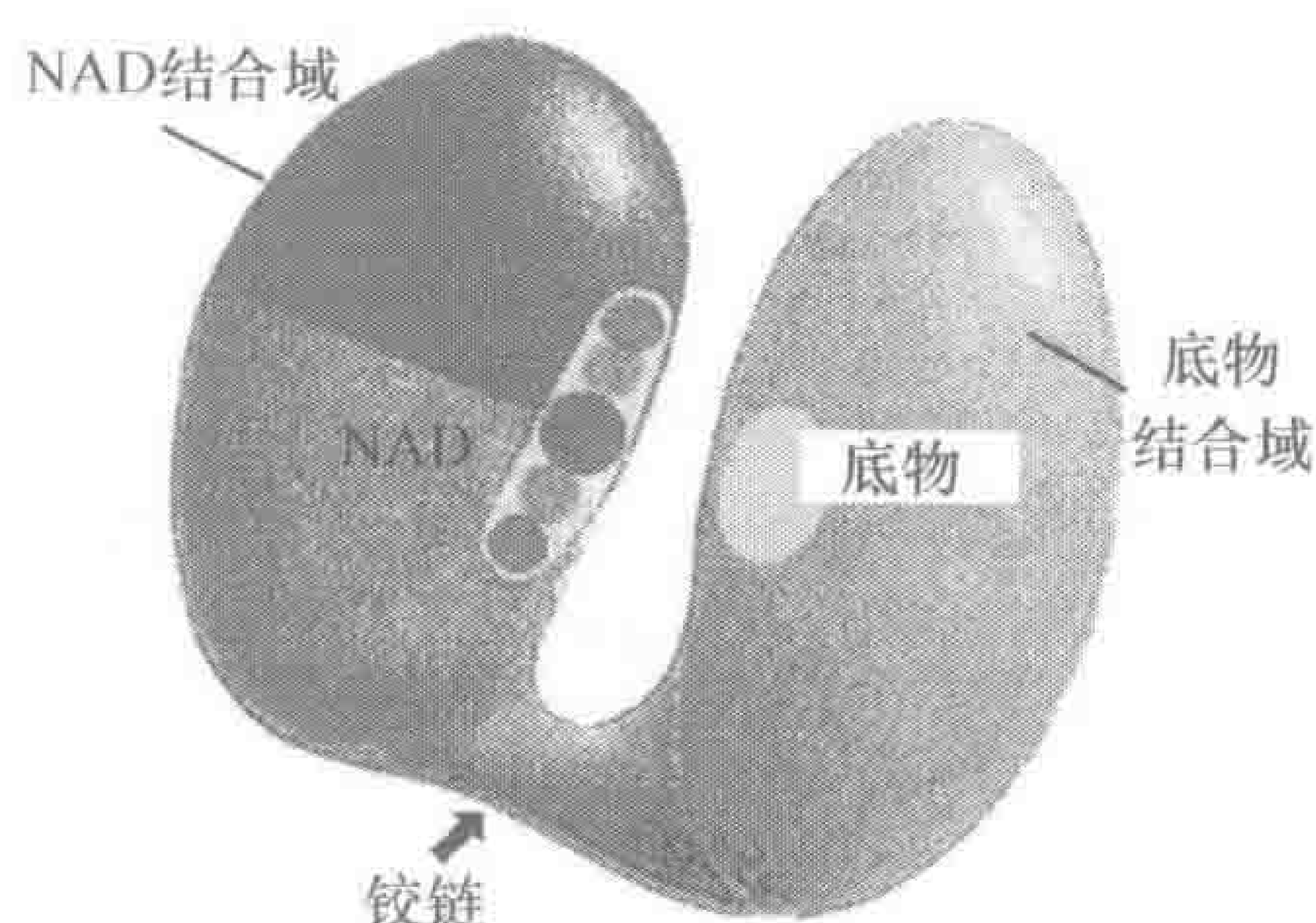


图 25-5　NAD 依赖性脱氢酶的肽链折叠成由相对柔软的铰链连接的两个独立的域。一个域基本是通用的(即很相似的域可见于不同的蛋白质中),它与辅因子 NAD 相结合。另一个域是非通用的,在每一种脱氢酶中都是单独不同的域,它结合所处理的底物。图片经许可取自文献[5]

脱氢酶由通过铰链连接的两个域构成(图 25-5)。

一个是结合所处理的底物的域,在每个脱氢酶中都是单独不同的构造。例如,在乙醇脱氢酶它含有 β 圆筒,而在甘油三磷酸脱氢酶中含有扁平 β 片。

与此同时,第二个域是与辅因子 NAD 相结合的,它们在所有的 NAD 依赖性脱氢酶中实际上是相同的,虽然在一些脱氢酶中它位于链的 N 端部分而在另一些中位于 C 端部分,而且尽管在所有的 NAD 结合域中的一级结构没有明显的同源性。在这个域(α/β 型)中,肽链具有"Rossman 折叠",而且相似性还分布到了各种脱氢酶的 NAD 结合域中的微小细节上:这些域的大部分能以 2Å 的精度相互比对,其中还包括 NAD 结合位点(这似乎表明,空间结构比一级结构能更好地记住蛋白质域的亲属关系)。

因此,不含催化中心的 NAD 结合域是通用的,脱氢酶行为的多样性是由分布

在不同构造的底物结合域上的底物结合中心的多样性所提供的：在整个蛋白质的结构中这两部分活性中心，靠近（图 25-5），形成一个活性中心，而脱氢酶的功能就是由这个活性中心提供的。

■ 接下来我想来讲述活性中心的非接触式相互作用，或者说，**变构**相互作用。在各种结合中心和活性中心之间的变构相互作用，尤其是在寡聚蛋白中的变构相互作用，在生化反应的控制和整合方面发挥着极其重要的作用。一个活性中心的状态"信号"通过于蛋白质球的形变（该形变能影响"另一个"活性中心）能被传递到另一个活性中心。

现在来讲述（以非常简化的形式）研究最充分的一种变构蛋白质：血红蛋白。

血红蛋白，确切地说，血红蛋白四聚体，是由两条 α 链和两条与之类似的 β 链组成的复合体［图 25-6(a)］。它的工作是在肺中（这里氧气很多）**结合**氧气，**转运**到肌肉（这里氧气很少）并**传递**给肌红蛋白，与血红蛋白四个亚基的任何一个都相似的单体蛋白。肌红蛋白的活性中心和血红蛋白的每个 α 亚基和 β 亚基都是血红素（或称为亚铁原卟啉）：一种辅因子，含有一个能结合一个 O_2 分子的铁离子［图 25-6(b)］。血红蛋白能够成功地运载 O_2 正是借助于变构作用，在本例中就是血红蛋白里面的四个血红素协调一致的相互作用。

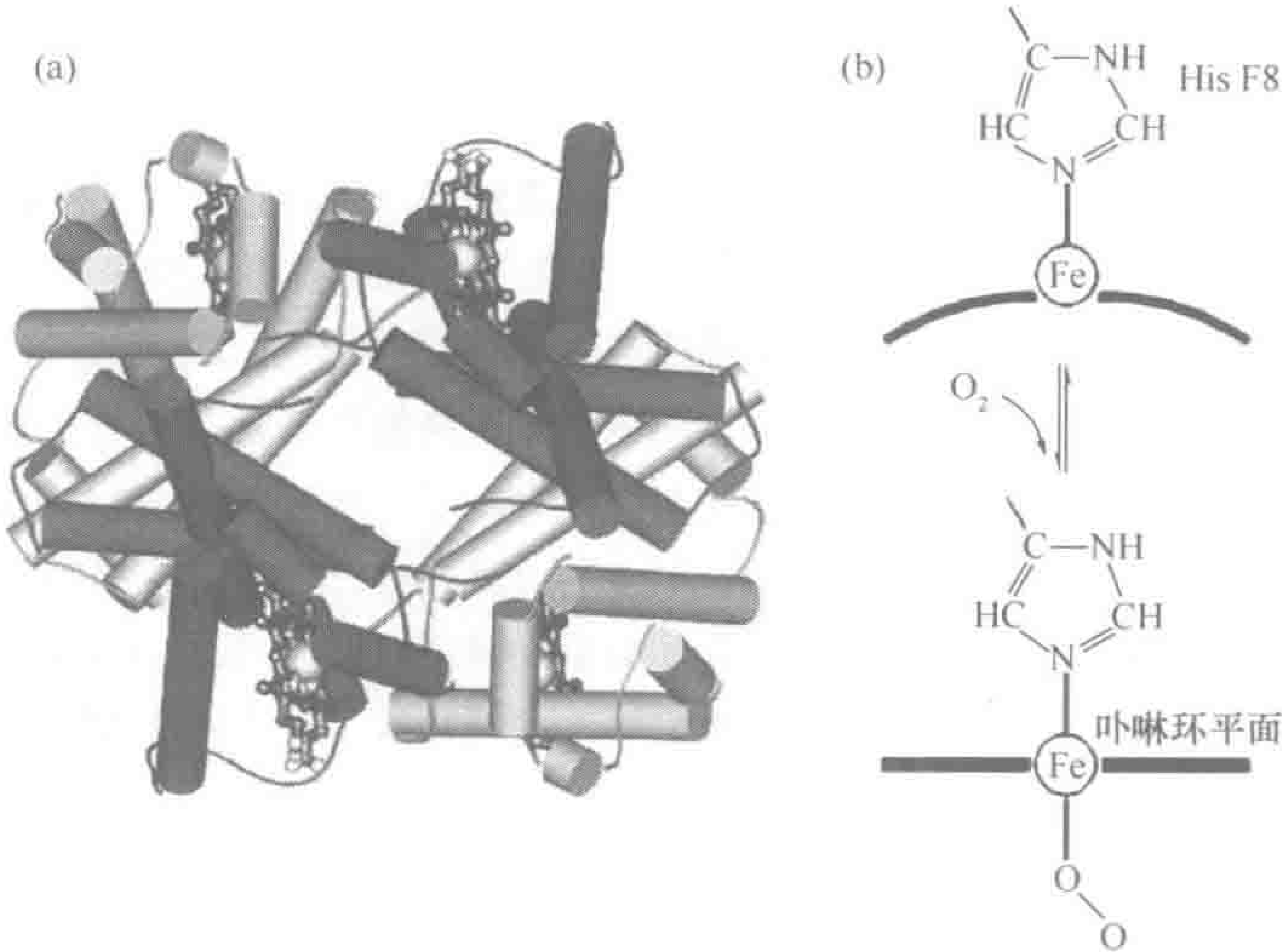


图 25-6 (a)血红蛋白的基本结构，由两条 α 链和两条 β 链组成。各条链以不同颜色显示；α 螺旋绘制为圆柱；血红素显示为内部含有铁离子（小球）的骨架模型。(b)血红蛋白中氧与血红素结合示意图。在与 O_2 结合之前 Fe^{2+} 处于高自旋态，有两个外层电子在两个不同的轨道上。因此 Fe^{2+} 的直径有点大，无法进入血红素的卟啉环中（它这时处于稍微弯曲的形状）。结合了 O_2 之后，Fe^{2+} 获得了低自旋的稍微更紧凑的形态，其中的两个电子自旋相反，占据同一轨道。这时 Fe^{2+} 可以进入卟啉环（这时它变成了扁平形状）。图中还显示了与铁离子形成配位键的 α 螺旋 F 中的组氨酸（HIS）。图片 b 取自文献［9］，略有修改

血红素中的铁离子由四个配体键 N—Fe^{2+}所支撑。在血红蛋白或肌红蛋白中，它还有一个与最靠近的组氨酸(α 螺旋 F 组氨酸)之间的配位键。它还可以与 O_2 分子连接形成最后一个配位键(或者与 CO 连接:这就是一氧化碳中毒时所出现的……)。

然而，这里有一个障碍:对血红蛋白和肌红蛋白的静态结构分析表明，O_2 分子无论如何也不能从外周进入到血红蛋白！事实证明，对此必不可少的是遮盖住血红素的几个侧基团的构象有协调一致的热运动。

当氧靠近血红素并与它的铁原子结合时，血红素的铁离子 Fe^{2+}的自旋状态发生改变，电子云的形状也因此发生改变。它的直径稍微变小，进入血红素卟啉环的中心。于是它(Fe^{2+})稍稍移动(约 1/2Å)，并把与它配位的 α 螺旋 F 组氨酸拉向自己。

因此，通过**电子-构象相互作用**(先发生快速的电子变化，然后发生构象变化)启动了血红蛋白一连串的构象形变。

组氨酸的继续移动，通过蛋白质的大量微小变形，使已结合了 O_2 的亚基的外形略有变化。它与其他亚基的接触被削弱。于是那些尚未结合 O_2 的亚基开始按顺序地松动(获得与蛋白质固有的、未受邻近蛋白质相互作用影响而变形的相同形状，以及相同的内部电子位移)，使它们的组氨酸开始把相应的铁离子推入到血红素的中心。此时这些铁离子已经能够容易地结合氧气了。结果是，一个 O_2 分子与血红蛋白的结合诱发与另外三个 O_2 分子的结合。同样如此，丢失一个 O_2 分子会导致全部丢失。

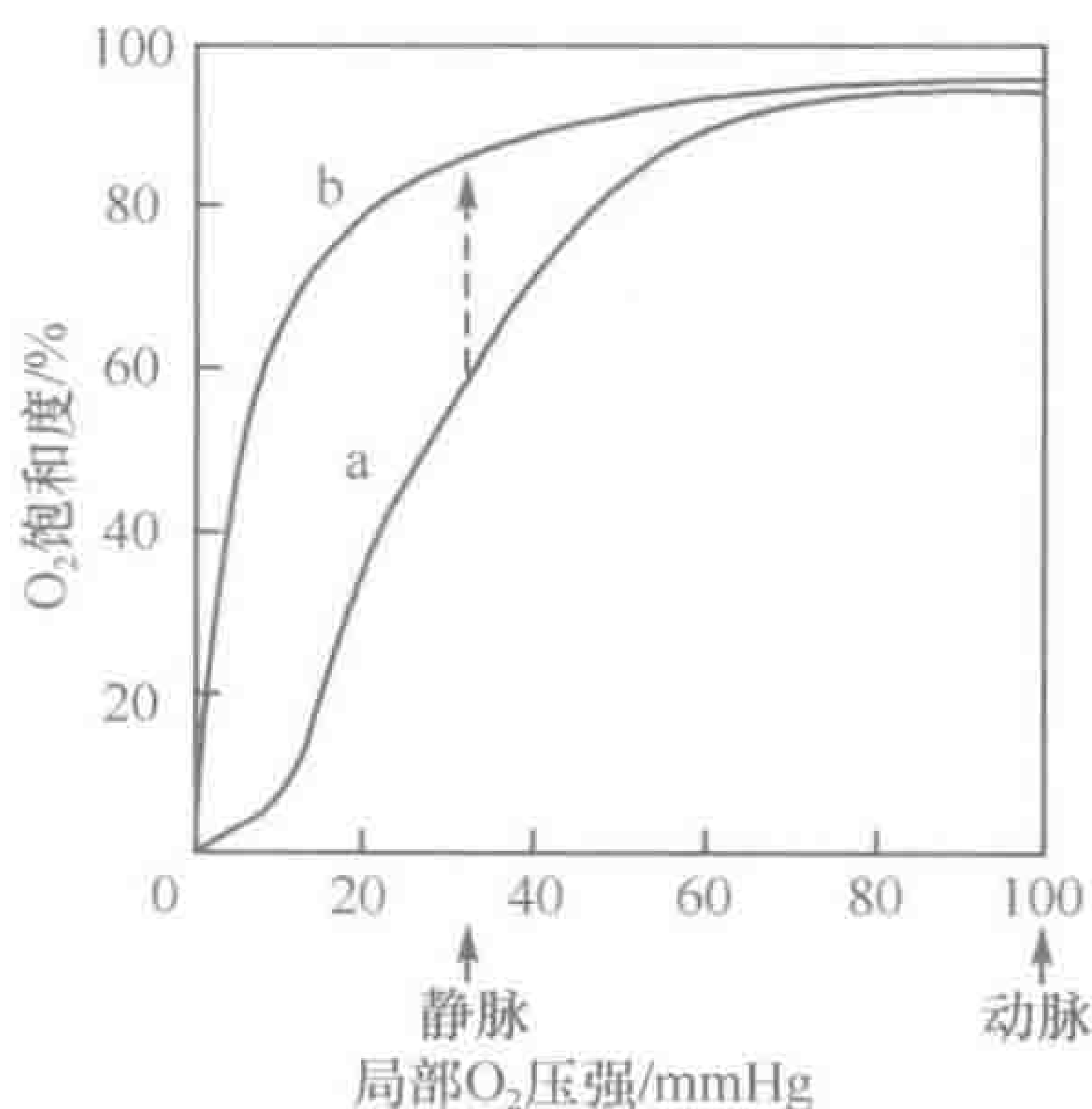


图 25-7 四聚体蛋白质血红蛋白 a 和单体肌红蛋白 b 的氧饱和度曲线。注意在氧气动脉压下 O_2 在肺中被血液里的血红蛋白所吸收;在氧气静脉压下 O_2 被血红蛋白传递给肌肉的肌红蛋白。O_2 与肌红蛋白结合的实验曲线对应于一级反应，而 O_2 与血红蛋白结合的 S 形实验曲线大约对应着三级(但不是正文中为简便起见而说的四级)反应。箭头对应着氧气从血红蛋白到肌红蛋白的转移。图片取自文献[1]，略有修改。1mmHg = 133. 322Pa

在血红蛋白中，伴随着氧气的结合和释放同时进行着两个重要的过程，也与其变构构象的重排有关。在未结合 O_2 的"脱氧"形式下，血红蛋白的亚基结合 CO_2(主要是以 HCO_3^- 的形式，但不是用血红素，而是用链的 N 端)和 H^+(用在空间上与链的 N 端接近的 C 端螺旋的组氨酸)。而在当血红蛋白结合在吸气中获得的 O_2 之后的"氧化状态"中，既丢失 H^+(这被称为"玻尔效应")——它与从 HCO_3^- 分裂出的 OH^-相结合之后形成 H_2O 分子，也丢失 CO_2 分子:都供呼气时排出。

从以上所述可以推论，血红蛋白可以认为是一种**同时**结合四个 O_2 分子的蛋白

质，而这也在反映在该（四聚体！）蛋白的氧饱和度与 O_2 浓度的非线性 S 形关系曲线上（图 25-7）。而肌红蛋白（单体！）的氧饱和度与 O_2 浓度的关系曲线**没有**这样的 S 形弯曲！

因此，在含有大量氧气的肺部中，血红蛋白是氧饱和的，而在组织中，O_2 的（静脉）压力相对较小，四聚体的血红蛋白**已经**释放 O_2，而单体的肌红蛋白**仍然**吸收 O_2，并进而通过肌肉（如那些使肺扩张和收缩的肌肉）氧化链传递 O_2。

■ 顺便谈一谈肌肉。我想以叙述它们的工作（或者确切地说，描述蛋白质的**机械化学**功能，即肌肉工作的基础）来开始结束这一讲。

在肌肉收缩［图 25-8（a）］中发挥着中心作用的是肌球蛋白［图 25-8（b）］。我想介绍快速横断条纹骨骼肌中的“肌球蛋白 II”的工作。其他类型的肌球蛋白都以类似的方式和类似的机械-化学机制工作，尽管在 kinezine、denein 和其他收缩蛋白质上观察到在细节上有差别。

肌球蛋白是一种二聚体蛋白。单体是由纤维部分，即“尾巴”（长长的 α 螺旋），和球状部分的“头部”组成。螺旋尾巴成对地缠绕成肌球蛋白分子，并以这种形式缠结成厚厚的肌纤维。长圆形的球状头部以约 150Å 的距离周期性地从这种纤维中伸出［图 25-8（c）］。它们可以与肌动蛋白丝［由球状肌动蛋白复合而成的原纤维，图 25-8（d）］相结合。通过“燃烧”ATP，这些“头部”能够沿着肌动蛋白丝“行走”，从而导肌肉收缩。

这种“行走”是如何发生的？ATP 化学势能是如何转化为机械运动的？

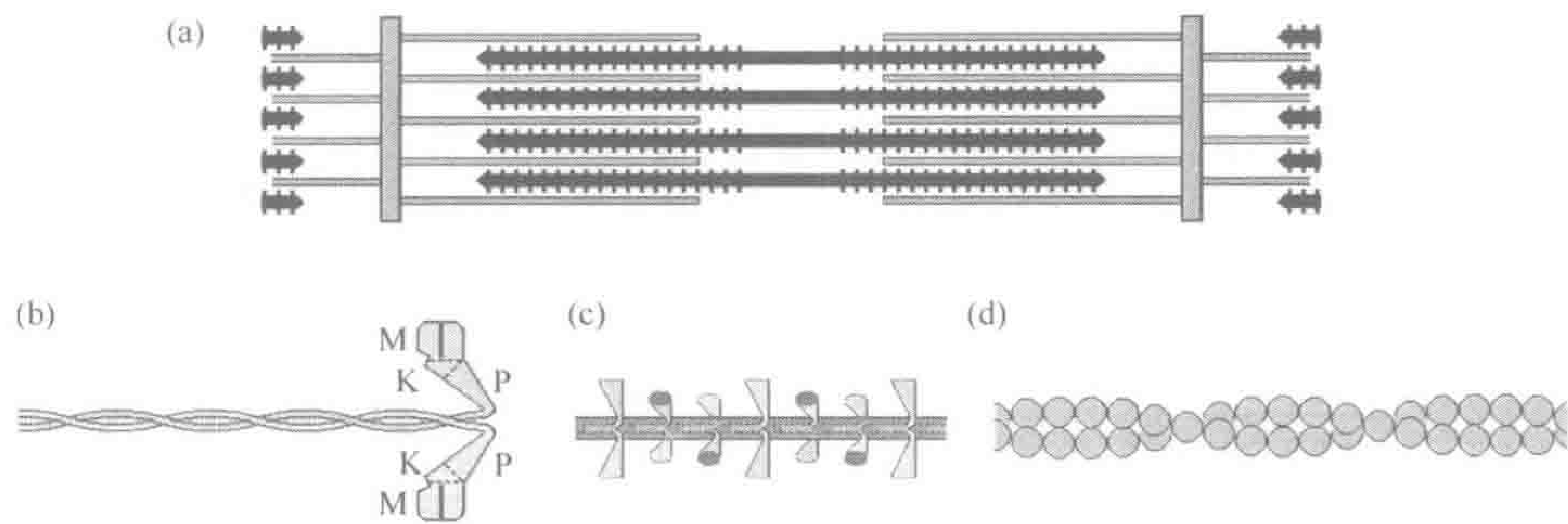


图 25-8 （a）肌节：肌肉的结构单元的示意图。粗肌球蛋白丝（含有从中伸出的头部）位于细肌动蛋白丝之间，后者被连接到固定肌节的基板上。肌动蛋白丝的长度约为 1μm。（b）由两个多肽链组成的肌球蛋白分子示意图。该分子中有两个头部和一条尾巴，是两个 α 结构的超螺旋。ATP/ADP 结合中心简略地表示为肌球蛋白头部中的小裂隙，其实是三个域组成（“马达域”M、“转换域”K 和“杠杆”P）。通过尾部缠结，肌球蛋白分子形成一条具有周期性在各个方向伸出头部的肌球蛋白丝（c），在本图中高度简化。因此每个双头的肌球蛋白分子沿着丝的中心轴相对于前一个的距离为 143Å。（d）肌丝：由肌动蛋白球体［肌动蛋白丝的其他成分（tropnin，tropomiosin，ADP）未显示］组装而成的双螺旋（周期为 360Å）的示意图。肌动蛋白只有在钙浓度足够高的情况下才能够与肌球蛋白结合

肌肉收缩可以相当圆满地由“摆动横桥模型”所描述。

这一模型建立在多个实验室多年的研究基础之上,这些研究包括收缩肌肉的X射线散射高速摄像照片,肌肉的电子显微镜观察,肌动蛋白头部在其不同状态下的X射线结构分析,单分子精细实验,以及其他许多研究。这些研究让我们看清楚肌肉收缩的大部分主要阶段。

研究表明,肌球蛋白的头部可以处于两种宏观构象的状态:直立态(与肌球蛋白丝呈90°)或“冲击态”(A)和静息态(R),即与肌球蛋白丝成倾斜角度≈45°。而且R态具有两种亚态:R′(缝隙打开,从催化中心朝向头部与肌动蛋白结合的区域)和R″,该缝隙关闭。这个缝隙既是ATP分解产物的出口,也能为头部与肌动蛋白结合的变构调节服务,在图25-9中有点夸大。从状态R转变到状态A是通过肌球蛋白头部的两个域(“杠杆”和“马达域”)围绕位于它们的“转换域”接头上的α螺旋转动来实现的。

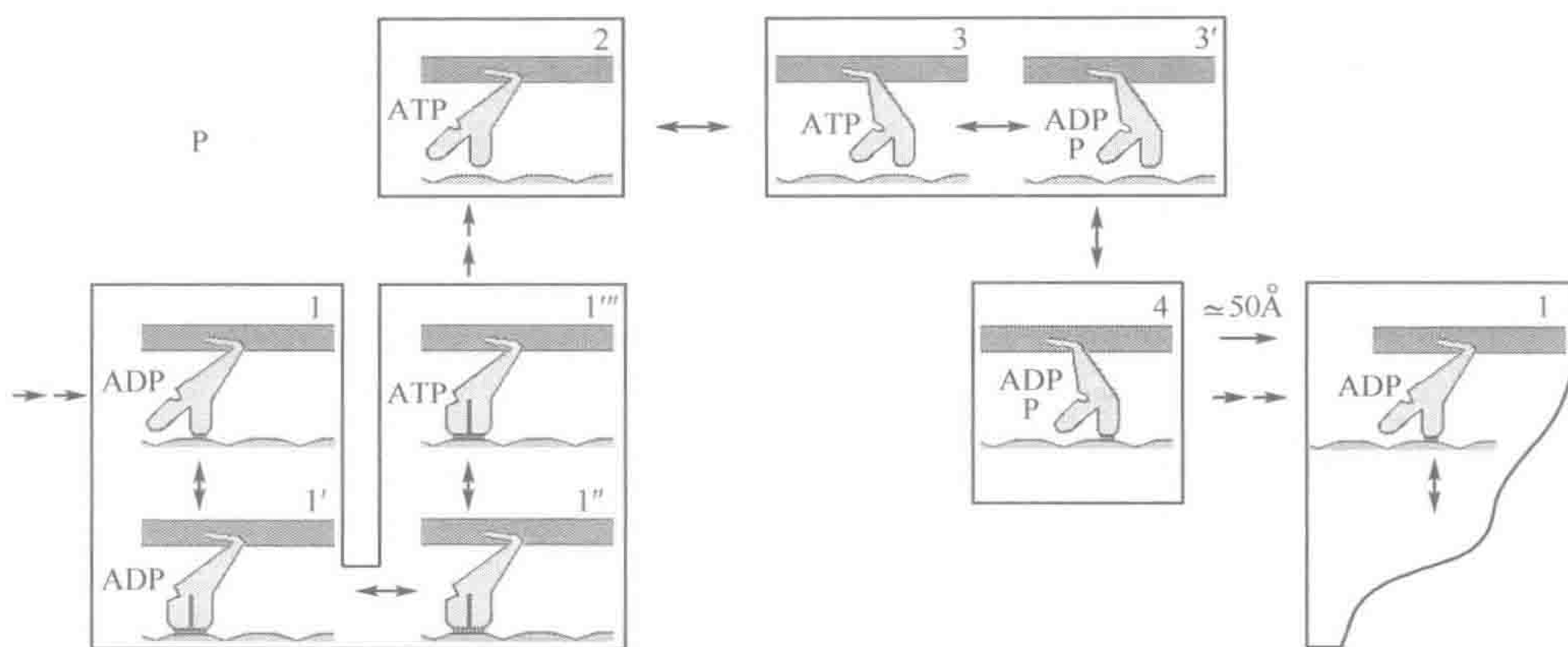


图25-9 肌球蛋白头部的机械化学循环。图中描绘了肌球蛋白的宏观状态和在该循环的不同阶段与之结合的小配体。肌球蛋白头部在阶段1′、1″和1‴(宏观状态R′,a+)与肌动蛋白紧密结合,在阶段4(A,a+)和1(R″,a+)轻微结合,而在阶段2(R″,a−)和3、3′(A,a−)不与肌动蛋白结合。在阶段1(R″,a+)、1′、1″、1‴(R′,a+)和2(R″,a−)头部是弯曲的,而在阶段3(A,a−)、3′(A,a−)和4(A,a+)是竖直的。在从阶段1‴转变到阶段2的过程中,在头部的马达域中缝隙张开(即构象R′转换为R″),而在从1转变到1′(从R″到R′)的过程中,缝隙关闭。在未结合的头部(a−)中缝隙是张开的。从阶段2到3和从4到1的转变中头部改变自身的弯曲度,而且从4转变到1(它与肌动蛋白的结合期间)伴随有“力量冲击”,导致所有肌球蛋白头部相对于肌动蛋白大约50Å的位移。从一个阶段到另一个阶段的转变或者是按宏观循环的转变(与肌球蛋白的构象变化相关,或与它和肌动蛋白的结合状态变化相关),或者是按微观循环的转变(与肌球蛋白和小配体(即ATP或其水解产物ADP和P)的结合状态变化有关)。从状态4到状态1的转变是由两个基本动作(一个是按宏观循环,另一个是按微观循环,双箭头标示)构成的;这两个动作的顺序尚未完全弄清楚:磷酸根(P)从肌球蛋白的排出可能是在力量冲击之前或紧随其后。在也以双箭头标示的1‴→2的转变中,头部上缝隙的张开可能是在从肌动蛋白脱离之前或紧随其后。那些在ATP、ADP和P的肌肉内浓度下实际上不可逆的转变以单向箭头表示,而那些容易可逆的用双向箭头表示(请注意,“力量冲击”和已结合ATP的头部从肌动蛋白上的脱离是不可逆的,而ATP的结合及其分解是可逆的)

此外,头部还能根据与肌动蛋白的结合处于两个宏观状态:结合态(a+)和非结合态(a-)。

肌球蛋白头部的这四个状态构成一个"宏观循环"(图 25-9):(R″,a+→R′,a+)→R″,a-→A,a-→A,a+→…请注意,在这个循环中的状态**超过两个**:超过两个状态才能建立定向循环(而两个状态只能是来回振荡),即产生**定向**运动。

此外,头部可以根据与小配体(ATP,ADP 和 P)的结合而处于四种状态:不结合,结合 ATP,结合两个 ATP 水解产物(ADP 和磷酸根 P),只结合一个 ADP。这四种状态构成一个"微观循环"。

"微观循环"与"宏观循环"相耦联(图 25-9):前者提供能量,而后者把能量转换为肌肉收缩。对这个耦联必需的是:"未经许可的"反应应该实际上被消除,以便"微观循环"的**每**一步**只能**在特定的"宏观循环"阶段中进行。

这样的耦联只能发生在这样的情况下:"头部"的构象变化和它与肌动蛋白结合性的变化能改变中心(该中心结合 ATP 或 ADP 和 P)的状态,而且,在特定的自身状态下,催化 ATP 到 ADP+P 的转化。而因为这个中心位于与肌动蛋白结合位点相距 40Å 处(与肌球蛋白头部的弯曲点所处位置大致相同),所以与肌动蛋白的结合变构性地引起 ATP/ADP 中心的变化:借助于连接这些中心的缝隙的构象变化(这与 O_2 与一个血红素的结合引起 O_2 与血红蛋白中其他血红素的结合常数的变化大致相同)。很自然,中心的这种连接也能以相反方向发生作用:肌球蛋白-ATP 复合体和肌球蛋白-ADP-P 复合体与肌动蛋白的结合比较弱,而肌球蛋白-ADP 复合体和"空闲"肌球蛋白头部与肌动蛋白的结合常数要超过 10 000~100 000 倍。

每次循环的过程中都要消耗一个 ATP 分子,这是在循环沿着引起肌肉收缩的方向进行的过程中,而不是在相反方向(在原则上,在后一种情况下,ATP 应该被合成,即使极其缓慢;然而对 ATP ⟶ADP+P 的逆反应需要把 ATP 的浓度从"正常的"毫摩尔降低到低于皮摩尔,而这会导致肌动蛋白丝的断裂!)。在 ATP 水解过程中获得的自由能[在对肌肉典型的 ATP($\approx 10^{-3}$mol/L)、ADP($\approx 10^{-5}$mol/L)和 P($\approx 10^{-3}$mol/L)浓度下约为 15kcal/mol ATP],被消耗在肌肉收缩上,巩固每一个之前的步骤。

长期以来,在对肌肉研究中的一个问题是,化学反应(ATP ⟶ADP+P)的能量是通过什么样的装置转化为肌肉拉力。因为化学活动本身(无论花费在事先准备上和等待能克服活化能垒的涨落上的时间有多少)占用的时间约为共价键振动的时间,即~10^{-14}s。而且参与到化学转变中的原子大约移动 1Å。而肌动蛋白丝在~10^{-3}s内移动"力量冲击"的长度($\approx$50Å)——这段时间比化学活动要长 1 亿倍! 在这时间里化学能量储存在哪里呢? 有一种假说,认为它储存在蛋白质的某些弹性成分中。然而对这些成分的寻找没有给出确定的结果:蛋白质的所有成分都具有某些弹性,但是找不到能提供 50Å 位移的弹性。然而,这样的弹性成分对肌肉的工作是必需的吗? 有人提出,对于"力量冲击",在从较不稳定(结合了 ADP)

的攻击构象(A)到更稳定(仍然是结合了ADP)的静息态(R)的转变过程中的杠杆本身的转动就足够了。这种转变是否占据整个"斜面下"的时间(就像在弹性应力的弛豫过程中),或者是否需要"无弹性地"越过某个能垒(在这样的情形下:能垒不高且在生物合理的时间内通过热涨落克服),都不要紧。

另一个问题。图25.9可能会让人产生一个印象,肌肉的工作需要,至少在"力量冲击"阶段,在所有彼此邻近分布并与肌动蛋白结合的肌球蛋白丝中有同步的结构转变。

然而,这种印象是不正确的。"工作中的"头部不是彼此接近地分布在肌球蛋白丝上。在每一个时刻,只有1%~2%的头部在"工作",推动肌动蛋白,而其他的在等待轮到自己。实验表明,在工作时的肌肉中每一个肌球蛋白的头部每秒钟只分解约20个ATP分子。然而在1/20s中,头部相对于肌动蛋白丝不是移动50Å(每步"力量冲击"的长度),而是3000~4000Å!这意味着,在这1/20s之外的98%~99%的时间里,头部不与肌动蛋白接触,而是在别人的肩膀上行走(而在为"力量冲击"准备的时候,经历显示于图25-9中的阶段2、3和3′)。然后,在少于1/1000s的时间内,它沿肌动蛋白带来自己的冲击,并继续在别人的肩膀上(更确切些,在"头部"上)行走……结果是,与肌动蛋白相结合的头部数目除以肌球蛋白丝数时,只有几个(存在于该肌动蛋白丝上的共有几百个),而悬挂在肌动蛋白质中的头部之间的距离有几千埃,这显然排除了它们的活动紧密同步的必要性。

头部的大尺度构象变化,和它与肌动蛋白的结合和不结合状态的轮换的耦联,让人想到在划船时的桨运动(也是一种循环:"桨在水后—桨在空气后—桨在空气前—桨在水前—桨在水后—……",如图25-9所示)。因此,肌球蛋白丝沿肌动蛋白的运动大致与桨沿水的运动相同(更确切些,是多桨的划艇,每一次划动中,一百个桨手中只有一个消耗自己的ATP分子)。

在结尾处,我还想向大家介绍一下驱动蛋白是如何工作的(我只介绍其中的一种:驱动蛋白1)。这种蛋白质通过燃烧ATP,能沿着微管运输mRNA、蛋白质复合体等,每分解一分子ATP就能从微管的"-"端朝"+"端行走8nm长的一步。驱动蛋白与肌动蛋白一样,连接着"宏观循环"和"微观循环",前者是其"头部"(实际上是作为脚)移动的机械循环,后者是分解ATP的化学循环。在"微观循环"过程中,驱动蛋白的头部("马达域")可能是与ATP结合,可能是与ADP结合,也可能是"空闲的"。在"宏观循环"过程中,驱动蛋白的头部可能有4种位置。首先,它可以在"后面"(即位于二聚体柄的后面,靠近微管的"-"端),也可以在"前面"(即位于二聚体柄的前面,靠近微管的"+"端)。其次,头部的这两个位置中的每一个都可能与微管蛋白结合或不结合。于是,由于头部与连接杆(相对于头部位置不同)之间的相互作用不同,"后面的"和"前面的"头部具有稍微不同的构象,并因此而具有不同的性质。实验已确认,对于驱动蛋白沿微管单

方向运动所必需的条件是：①ATP **只能**由“后面”的头部所分解；②ADP **只能**从“前面”的头部所释放；③磷酸根所占据的头部（无论是位于“前面”还是“后面”）与微管蛋白只能微弱结合；④“空闲”的头部，以及结合了ATP的头部，与微管蛋白强烈结合并实际上不能与之脱离。于是驱动蛋白在所有时间内总是至少有一个头部与微管结合，而另一个头部可以扩散性地从ADP不能脱离的位置移动到ADP可以脱离的另一个位置。

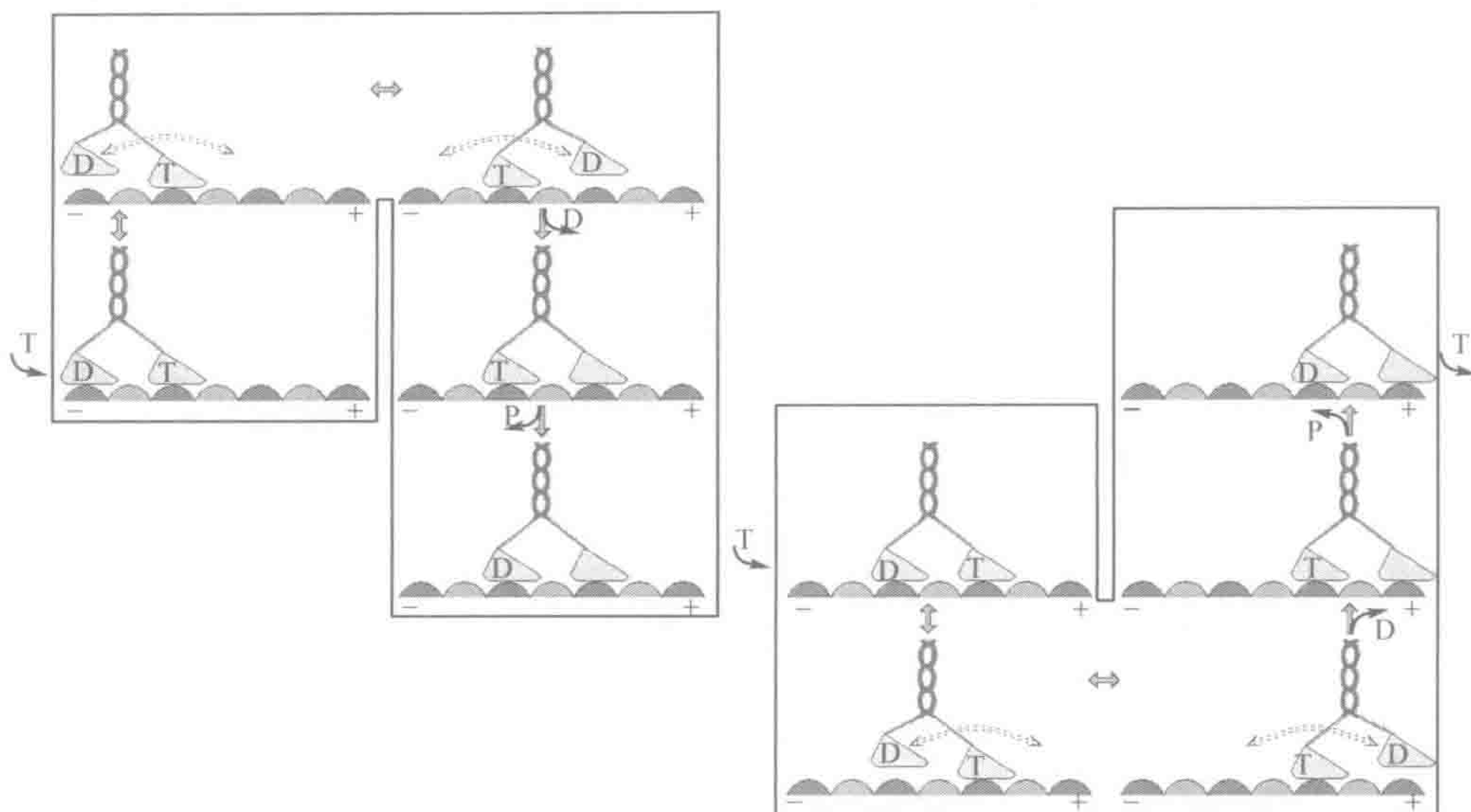


图25-10　二聚体驱动蛋白1沿微管运动示意图。微管蛋白球体（半径4nm）外层显示为半球面。两种不同颜色的圆角三角形表示驱动蛋白的两个头部，双螺旋表示它们共同的柄，直线表示连接头部与柄的连接段。驱动蛋白的两个头部是相同的，以相同的方向（相对于微管蛋白）粘接在微管上。图中显示了驱动蛋白移动的完整动作，这是由两步组成的，每一步都把“宏观循环”和“微观循环”连接在一起。每个亮灰色的箭头表示驱动蛋白头部运动的基本动作。长的灰色虚箭头表示头部在正负16nm的移动[在此过程中驱动蛋白的柄（这个二聚体蛋白利用它进行整体移动是很方便的）移动8nm]。黑箭头表示底物（T＝ATP，D＝ADP，P＝磷酸根）状态变化的一个基本动作。在ATP、ADP和磷酸根的细胞内浓度下实际上不可逆的转变以单向箭头表示，容易可逆的转变以双向箭头表示。本图片根据文献Yildiz，Tomishige，Gennerich.，et al.2008.Cell 134：1030－1041拼制而成

结果是，驱动蛋白沿着微管行走，肌球蛋白沿着肌动蛋白行走，肌肉收缩，肺呼吸空气，氧气与血红蛋白结合并被传递给肌红蛋白，肌红蛋白把它传送给肌肉和其他器官，我们就这样呼吸和生活。

后　记

令我深为遗憾的是,我不得不独自完成这些篇章:我的老师,Oleg B. Ptitsyn,逝世了……因此本讲义中的所有不足之处由我一人承担。我将感谢把本书不足之处告知我的所有读者。

A. V. 芬克尔施泰因

2000 年

afinkel@ vega. protres. ru

译后记

2002 ~2006 年期间，我作为国家公派留学生在俄罗斯莫斯科大学生物系学习期间，有幸聆听了 Alexei V. Finkelstein 教授为三年级大学生们讲授的《蛋白质物理》课程。这门课程中旁征博引的丰富信息和精彩纷呈的学术思想给了我很多的启迪，而含有强烈个人风格的教材也让我深为赞叹，因而决心把它译为中文介绍给我国的读者。翻译工作是从 2008 年开始的，历经多年才得以完成。衷心希望读者能喜爱本书。

本书的出版得到了华中师范大学物理学院贾亚教授和华中科技大学物理学院肖奕教授的支持和帮助，在此表示诚挚的感谢！

由于本书具有明显的跨学科性质，书中的专有名词和专业术语很多，虽然已作了最大努力，但是由于译者水平有限，错误和不当之处在所难免，敬请广大读者批评指正。

任何批评和建议都可以发送给如下电子邮箱：lianbang@ phy. ccnu. edu. cn。

本书的补充材料和重要的读者意见都将列于如下网站：http://bio. phy. ccnu. edu.cn/proteinphy。

李安邦

2013 年 6 月

推荐文献

(这些文献同时也是本书中很多插图的来源)

[1] Фершт Э.Структура и механизм действия ферментов,-М:Мир,1980.гл.1,8-12.
Fersht A.Enzyme structure and mechanism,2nd ed.-NY:W.H.Freeman & Co.,1985.Ch.1,8-12
Fersht A.Structure and mechanism in protein science:A guide to enzyme catalysis and protein folding,.-N.Y.:W.H.Freeman &Co.,1999.

[2] Волькенштейн М.В.Биофизика.-М:Наука,1981; гл.4,6.
Volkenshtein M.B.生物物理学.1981.(只有俄文版)

[3] Шульц Г.Е.,Ширмер Р.Х.Принципы структурной организации белков.-М:Мир,1982.
(Schulz G.E.,Schirmer R.H.Principles of protein structure.-N.Y.-Heidelberg-Berlin:Springer-Verlag,1979).

[4] Кантор Ч.,Шиммель П.Биофизическая химия,М:Мир,1982.т.1,гл.2,5; т.3,гл.17,20,21.
Cantor C.R.,Schimmel P.R.Biophysical chemistry.-San Francisco,W.H.Freeman & Co,1980.Part 1,Ch.2,5; Part 3,Ch.17,20,21.

[5] Branden C.,Tooze J.Introduction to Protein Structure.-N.Y.,London:Garland Publ.,Inc.,1991.2nd ed.,1999

[6] Creighton T.E.Proteins,2-nd ed.-N.Y.:W.H.Freeman & Co.,1991.

[7] Perutz M.F.Protein structure.-N.Y.:W.H.Freeman & Co.,1992.

[8] Ленинджер А.Основы биохимии.В 3-х т.-М.:Мир,1985.Гл.4-8,23,29.
Leninger A.L.,Nelson D.L.,Cox M.X.Principles of biochemistry.2nd ed.-N.Y.:Worth Publ.Inc.,1993.Ch.3-8,21,26.

[9] Страйер Л.Биохимия.В 3-х т.– М.:Мир,1984 (т.1):1985 (т.2,3).Гл.1-9,27,33-34.
Stryer L.-Biochemistry,4th ed..-N.Y.:W.H.Freeman & Co.,1995.Ch.1-3; 7-16; 34-36.

[10] Полинг Л.Общая химия.-М.,Мир,1974.Гл.1-6,9-13,16,24.
Poling L.General chemistry.-N.Y.:W.H.Freeman & Co.,1970.Ch.1-6,9-13,16,24.

[11] Степанов В.М.Молекулярная биология.Структуры и функции белков.-М.,Высшая Школа,1996.
(Stepanov V.M.分子生物学.蛋白质结构和功能.1996.只有俄文版)

[12] Чернавский Д.С.,Чернавская Н.М.Белок-машина.Биологические макромолекулярные конструкции.-М.:Изд-во МГУ,1999.
(Chernavskii D.C.,Chernavskii N.M.蛋白质——机器.生物大分子结构.1999.只有俄文版)

[13] Rubin A.B.Biophysics.1999.第一卷第7-14章.(俄文版和英文版)

[14] Эмануэль Н.М.,Кнорре Д.Г.Курс химической кинетики.4-е изд –М.:Высшая школа,1984
(Emanyel N.M.,Knorre D.G.化学热力学教程(第4版).1984.只有俄文版)

[15] Howard J.-Mechanics of motor proteins and the cytoskeleton.Part III.-Sunderland,Massachusetts:Sinauer Associates,Inc.,2001.

[16] Serdyuk I.N.,Zaccai N.,Zaccai J.Methods in molecular Biophysic.Structure,function,dynamics.-Cambridge,England:Cambridge University Press,2007

为了方便阅读,在正文中只显示这些文献的链接。而原始工作和综述直接在图片或表格的题注中给出。

蛋白质图片取自蛋白质结构数据库(Protein Data Bank,PDB)(下面这篇文献介绍了它的建立:Bernstein F.C., Koetzle T.F., Meyer E.F., Jr, Brice M.D., Kennard O., Shimanouchi T., Tasumi T., // J.Mol.Biol.1977, V.112.P.535-542),并用程序 MOLSCRIPT(Kraulis P.J., J.Appl.Cryst.1991.V.24;P.946-950)、WHAT IF (Vriend G.// J.Mol. Graphics.1990.V.8.P.52-56)、RasMol (Sayle R., Milnerwhite E.J.// Trends biochem.Sci.1999.V.20.P.374-376)、Insight II(Molecular simulations Inc., 1998)和 ViewerLite(Accelrys Inc., 2001)绘制。

练习、解答与评注

学习精密科学的要求之一是学习解答练习题。缺少练习题是本书以前版本的重大不足。因此,在本书的俄文第3版中,我尽力弥补这一空白,而在本版中,更增加了习题的数量。

因为本书预计为具有完全不同专业知识的人们所使用,有些习题显得极为复杂①,而另一些则非常基础。基本上,"非常基础"的习题是对读者(虽然是某个科学领域中的优秀专家,却在另一个领域里遇上了困难)来信的回应。由于这个缘故,我对习题给出了相当详细的解答,甚至对大部分读者会显得有点多余。

A. V. 芬克尔施泰因,2004-2011

第2讲

习题2.1

氨基酸有"左手型"和"右手型"。水是否有"左手型"和"右手型"? 乙醇呢? 为什么?

解答:

只有那些无法通过在三维空间中的移动和转动、或通过绕价键旋转而与自己的镜像所重合的分子,才有"左手型"和"右手型"。

水分子,H—O—H,总共由三个原子组成。三个点(H、O和H的原子核,我们忽略电子的空间构象)总是位于同一平面内,而平面物体的镜像总是可以通过在三维空间中的平面移动和翻转而与原始物体重合(拿任意三角板试试!)。因此"左手型"和"右手型"的水是完全不可能的。

原则上来上,"左手型"和"右手型"只有三维物体才具有。因此由四个或更多原子组成的分子(如CH_3CH_2OH)的瞬时构象可能与其镜像不同。但是,如果在这种镜像分子中把所有内部旋转角χ_i替换为χ_{-i},我们能达到初始的构象CH_3CH_2OH。因此CH_3CH_2OH的镜像("左手型"和"右手型"的乙醇)只有在热运动不能消除其对初始构象的记忆时才能达到,即在室温下$\sim 10^{-10}$s范围内(在超低温下有更长的时间);但是在长时间内无法观察到"左手型"和"右手型"

①有些习题实在是过于复杂和困难,中译本中省略未译出。

的 CH_3CH_2OH。

但是在“重”乙醇 $CH_3CDH—OH$（其中 D 为氘）具有“左手型”和“右手型”构形！此时中心原子 C 有四个**不在同一平面内**的共价键，而且所有原子（C，H，D，O）**各不相同**，而原子从一个键转变到另一个键需要很大的激活能，实际上需要无限长的时间。

而“重”乙醇 $CDH_2CH_2—OH$ 不具有“左手型”和“右手型”构形：它的中心原子 C 的四个共价键（与 C，H，H，O 相连）中的两个是相同的，第一个 C 原子（与 C，H，H，D 相连）也是一样，因此这种分子的“镜像”构形能通过绕共价键转动转变为初始构形（正如 CH_3CH_2OH 那样）。

习题 2.2

具有能量 kT 的热振动在 27℃ 时具有频率 $\nu_T \approx 7\times10^{12}\,s^{-1}$。在 −200℃ 时热振动的频率如何？在 +2700℃ 呢？

解答：

热运动的频率由公式 $kT \approx h\nu_T$ 所决定，其中 T 为绝对温度。27℃ 对应于绝对温度 $T = 273 + 27 = 300(K)$，−200℃ 对应于绝对温度 $T' = 273 - 200 = 73(K)$，即低到 1/4，+2700℃ 对应于绝对温度 $T'' = 273 + 2700 = 2973(K)$，即高达 10 倍。因此在 −200℃ 时热振动频率比 27℃ 时低得多，$\approx 1.7\times10^{12}\,s^{-1}$，而在 +2700℃ 时高达 27℃ 时的 10 倍，$\approx 7\times10^{13}\,s^{-1}$。

习题 2.3

C—H 键的振动频率 $\nu_{C—H}$ 为 $7\times10^{13}\,s^{-1}$。该振动与温度的关系如何？

解答：

振动频率与温度无关：它只取决于价键的刚性和所结合原子的质量。振动的可激活性取决于温度：当 $\nu_{C—H} > \nu_T$ 时，振动实际上并未激活；只有当 ν_T 升高到 $\nu_{C—H}$，即 ≈ +2700℃ 时（参见上一题），振动才能激活。

习题 2.4

求出对应于频率为 $\nu = 7\times10^{12}\,s^{-1}$ 的振动的光波的波长。

解答：

光波的波长 $\lambda_{光} = c/\nu$，其中 c 为光速（$3\times10^8\,m/s$）。则 $\lambda = (3\times10^8\,m/s)/(7\times10^{12}\,s^{-1}) \approx 0.4\times10^{-4}\,m = 40\mu m$。

习题 2.5

求出对应于频率为 $\nu = 7\times10^{12}\,s^{-1}$ 的振动的声波的波长。设声音在水中的速度为 $u = 1500\,m/s$。

解答：

由公式可得 $\lambda_{声} = u/\nu = (1500\,m/s)/(7\times10^{12}\,s^{-1}) = 0.2\times10^{-9}\,m = 2Å$。但是由于其波长接近于原子间距离，所以这种声波衰减极快。

习题 2.6

电子绕原子核振动的圆频率为 $\omega=2\times10^{16}s^{-1}$。试根据海森堡不确定性原理估算该振动的(1)振幅和(2)动能。

解答:

根据海森堡不确定性原理,振幅的不确定量($\Delta p=m\Delta v$)和坐标的不确定量(Δx)满足关系式 $\Delta p\Delta x\approx\hbar$。其中 $\hbar=1.05\times10^{-34}J\cdot s$("动量矩量子"),由公式 $\hbar=h/2\pi$ 与普朗克常量($h=6.6\times10^{-34}J\cdot s$)相联系。

以振幅 A 振动时,坐标在 $+A$ 至 $-A$ 间变化,即其不确定量 $\Delta x=[A-(-A)]/2=A$。而在圆频率 ω 时速度在 $+\omega A$ 至 $-\omega A$ 间变化,即其不确定量为 $\Delta\nu=\omega A$。因此 $\Delta p\Delta x=mA^2\omega\approx\hbar$,由此可得:振幅 $A\approx(\hbar/m\omega)^{1/2}$,而动能 $E=m(\omega A)^2/2\approx\hbar\omega/2$。

由于电子质量为 $m=0.91\times10^{-30}kg$,圆频率为 $\omega=2\times10^{16}s^{-1}$,可得 $A\approx0.75\times10^{-10}m=0.75Å$,$E\approx1\times10^{-18}J\approx0.25\times10^{-18}cal/mol=150kcal/mol$。

习题 2.7

在 C—H 键中的质子的谐振动频率为 $\nu=7\times10^{13}s^{-1}$。试根据海森堡不确定性原理,估算其振幅。

解答:

频率 ν 与圆频率 ω 之间的关系为 $\nu=\omega/2\pi$,即在 $\nu=7\times10^{13}s^{-1}$时有 $\omega=4.4\times10^{14}s^{-1}$。由习题2.6中所得的公式:振幅 $A\approx(\hbar/m\omega)^{1/2}$,得到,在质子质量为 $m=1.67\times10^{-27}kg$ 和频率 $\nu=7\times10^{13}s^{-1}$时,振幅 $A\approx0.12Å$。

习题 2.8

估算热振动中的粒子的坐标量子不确定量。粒子的分子质量分别为:(1)1Da①(H 原子);(2)18Da(水分子);(3)100Da(氨基酸残基)。

解答:

根据海森堡不确定性原理,$m\Delta v\Delta x\approx\hbar$,其中 Δv 为粒子速度的量子不确定量,Δx 为其坐标的不确定量。在振动中有 $\Delta v\approx\omega\Delta x$(见习题2.6),其中 ω 为振动圆频率,于是 $\Delta x\approx(\hbar/m\omega)^{1/2}$。热振动的特征圆频率由公式 $\hbar\omega_T\approx kT$ 所决定。因此热振动中的粒子的坐标的量子不确定量为 $\Delta x_T\approx(\hbar/m\omega_T)^{1/2}=\hbar/(mkT)^{1/2}$.

在 $T=300K$(即 $kT=2.5\times10^3J/mol=2.5\times10^3\ J/(0.6\times10^{24})$)且粒子质量 $m=1Da=1.67\times10^{-27}kg$ 时,有 $\Delta x_T\approx0.4Å$;在 $m=18Da=18\times(1.67\times10^{-27}kg)$ 时,$\Delta x_T\approx0.09Å$;在 $m=100Da$ 时,$\Delta x_T\approx0.04Å$。

第3讲

习题 3.1

(相当复杂,略)

①关于分子质量单位 Da,参见第10讲中的脚注。

习题 3.2

范德华相互作用取决于围绕原子核振动的电子的相互极化。问题：范德华相互作用依赖于原子的内部电子还是外部电子？

解答：

依赖于外部的电子：它们与核的联系微弱，因此容易极化。

习题 3.3

一对电子的范德华相互作用与其周围的媒质有关吗？

解答：

是的。因为成对原子的范德华相互作用取决于这些原子的极化状态的电偶极子-电偶极子相互作用的平方，而电偶极子-电偶极子相互作用与任意静电相互作用一样，都反比于媒质的介电系数。因此成对原子的范德华相互作用随着媒质介电系数的升高而降低。当然这里所说的介电系数对应于电子的本征振动频率，即电磁辐射谱的紫外区域。这一介电系数取决于原子的电极化率和单位体积内的原子数，即媒质密度。液体、晶体和玻璃的典型介电系数约为 2。[注意，在密媒质中单独取出的一对电子的范德华相互作用的减弱(与其在真空中的相互作用相比)并不大，根据对单原子液体和固体的计算为 ~5%，因此在密媒质中的减弱效应通常可以忽略。然而可以证明(Finkelstein A. V. 2007.Chemistry Central J. 1:21)，在存在有共价键，即在多原子分子的媒质中，这种效应会大大增强。]

习题 3.4

(相当复杂，略)

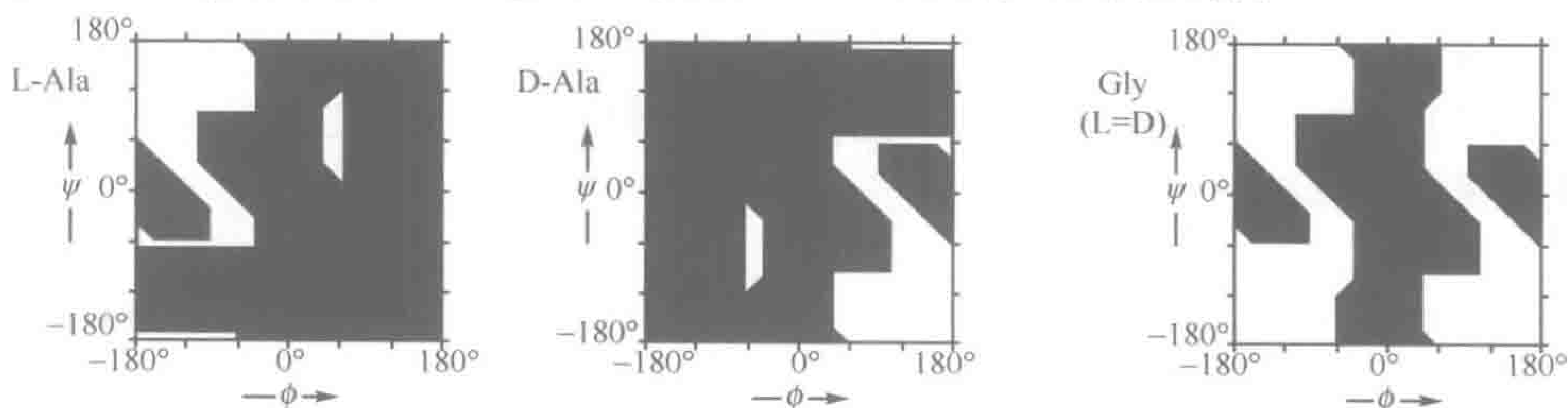
习题 3.5

回顾 L 型 Ala(丙氨酸)和 Gly(甘氨酸)的允许状态图。这两个氨基酸的 D 型镜像异构体的允许状态图会是什么样？

解答：

D 型 Ala 是 L 型 Ala 的镜像，后者的允许状态图显示于图 3-4 中。在镜像中，绕轴的旋转角(尤其是 ϕ 角和 ψ 角)相反地改变自己的符号(图 2-3)。因此 D 型 Ala 允许状态图(见下图中的白色区域)显得像是 L 型 Ala 的允许状态图转动 180°的结果。

D 型 Gly 和 L 型 Gly 就是同一种甘氨酸，—NH—CH_2—CO，是自身镜像对称的。相应地，其允许状态图(图 3-3)转动 180°之后仍与自身重合。

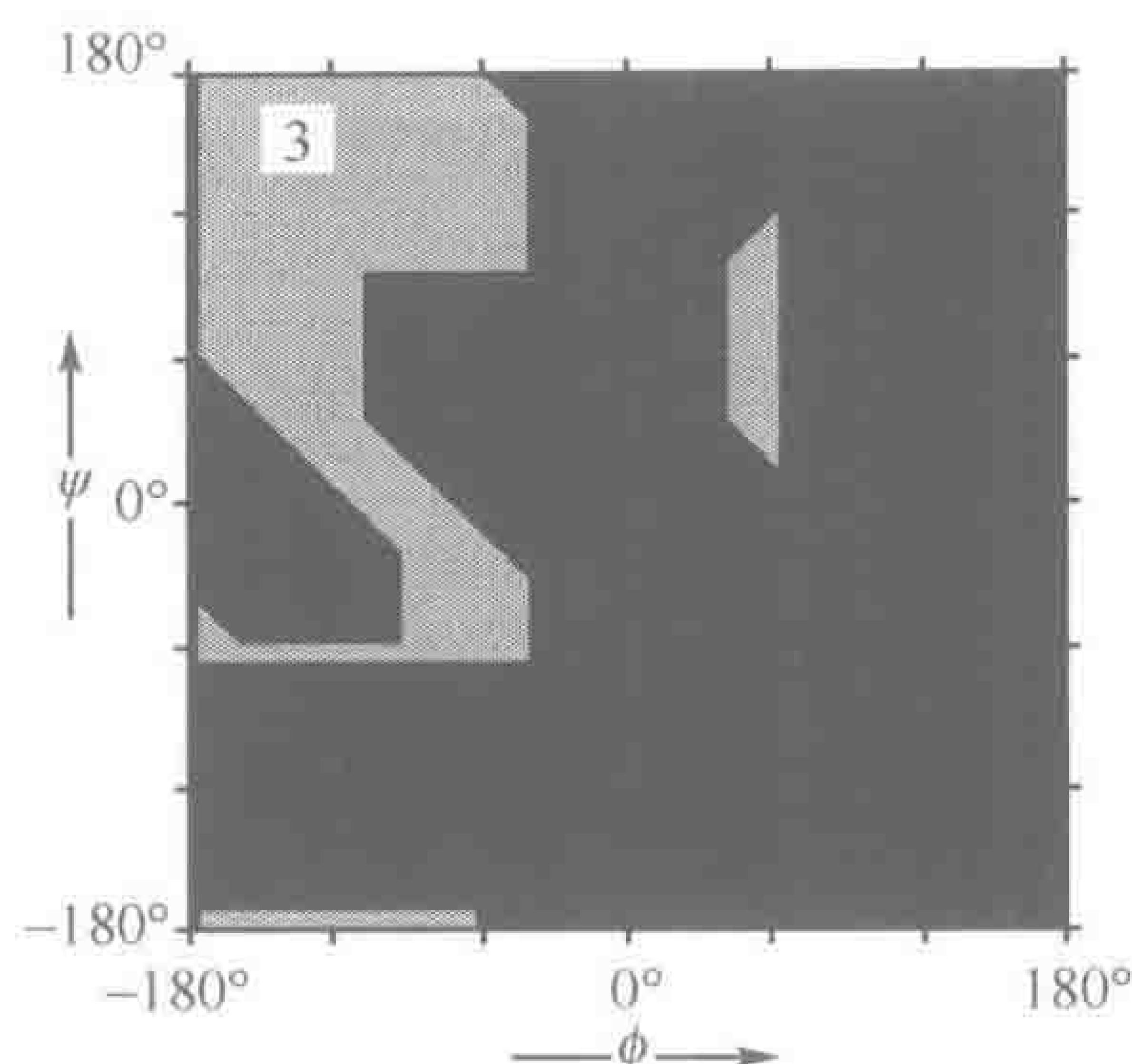


习题 3.6

假设在多肽链中的 L 型氨基酸残基带有三个 C^{γ} 原子。该残基的允许状态图如何?

解答:

图 3-5 表明,只有在一个区域中 ϕ 角和 ψ 角同时对 C^{γ}原子沿 χ_1 角转动的所有三个旋转异构体都允许。只有在这个区域才对带三个 C^{γ}原子的残基是允许的(见下图中的白色区域)。



第 4 讲

习题 4.1

已知,在 O—H∷∶O 键中 O 原子中心之间的距离等于 2.8Å,O—H 共价键的长度等于 0.96Å,而在晶体中观察到的 O 原子间最小距离等于 2.7Å。试估算在水分子中,H 原子中对于其他原子无法穿透的“固体”部分的半径。

解答:

O 原子的“固体”部分的半径等于 2.7Å/2 = 1.35Å。因此在水分子中,H 原子中对于邻近的 O 原子无法穿透的“固体”部分等于 2.8Å−2.7Å/2−0.96Å = 0.49Å。它只有在非极化原子中 H 原子的“正常”最小可能半径($r_{\min}$)(显示于表 3-1 中)的 1/2。这是 H 电子被负电性的 O 原子拉开和剩余在 H 中的电子云被 H^+与 O^-的静电相互作用“修饰”的结果。

习题 4.2

有一些关于“狗洞”的传说,小狗会闷死在那种山洞里,而人可以通过。这种现象通常是这样解释的:洞里分泌出的二氧化碳(CO_2)会沉在洞底,从洞底分泌出的较轻的氮气和氧气形成“路边”的空气层,而只有这种空气层是适合于呼吸的。这种解释合理吗?

解答:

不完全合理。如果全部原因仅取决于单个分子的话,CO_2 实际上在整个山洞的空间里具有几乎相同的浓度。而实际上,根据气压关系式,概率 ~ exp($-mgh/$

kT)，重力作用产生 CO_2（分子质量为 44Da）分子层特征高度 $h^* = kT/mg$ 应该比 O_2（分子质量为 32Da）分子层的特征高度低，为后者的 $44/32 \approx 1.4$ 分之一。由于空气层的特征高度 h^* 为几英里，则 CO_2 层的特征高度也会是几英里……这意味着，"狗洞"的致命性不是由于 CO_2 按空间的平衡分布引起的，而是由 CO_2 在洞中的积累动力学引起的。在"狗洞"中，CO_2 要么是从底部分泌出来，要么就实际上就是由墙壁分泌出来并流到底部的：由于扩散缓慢（CO_2 流入到底部的速度相比）和无风而在底部长时间停留。然而，即使 CO_2 不再增加，扩散（即使无风）也会逐渐地把它传播到整个山洞，使它变得对呼吸不利。

第 5 讲

习题 5.1

假设某种分子是刚性的（即不具有可激发的内部自由度），并具有球体形状。假设，在 27℃时，它在晶体中振动时的振幅为 $a = 0.5$Å，而它在晶体上的饱和蒸汽气压为 $P = 0.1$ 标准大气压。试估算其汽化时的自由能变化量 ΔG，熵变 ΔS 和焓变 ΔH 。

解答：

$\Delta G = 0$，$\Delta S \approx 25.8\text{cal/(g} \cdot \text{mol)}$，$\Delta H \approx 7.7\text{kcal/mol}$

解释如下。

首先，因为根据定义，在分子由固体（或液体）迁移到饱和蒸汽中时的自由能变化等于 0，所以 $\Delta G = 0$。然后在分子从晶体（其振动的体积为 $V_{cr} = (2a)^3$）迁移到蒸汽中时（其体积变为 $V_g = RT/P$），熵变化为 $\Delta S = R\ln(V_g/V_{cr})$。最后，焓变化为 $\Delta H = \Delta G + T\Delta S$。

计算时需要把所有数值数据按统一量纲（如国际单位制）代入。

27℃为 $T = 300\text{K}$；此时有 $RT = 1.99\text{cal/(g} \cdot \text{mol)} \times 300\text{K} \approx 600\text{cal/mol} \times 4.18\text{J/cal} \approx 2.5 \times 10^3 \text{J/mol}$；而 1 标准大气压 ≈ 101kPa，则 0.1 标准大气压 ≈ 10kPa = $1.0 \times 10^4 \text{N/m}^2$。

于是有：

$V_{cr} = (2a)^3 = 1(\text{Å}^3)$；

$V_g = RT/P \approx (2.5 \times 10^3 \text{J/mol})/(1.0 \times 10^4 \text{N/m}^2) = 0.25\text{m}^3/\text{mol}$

$= 0.25\text{m}^3/(0.602 \times 10^{24}) = 0.42 \times 10^{-24}\text{m}^3 = 0.42 \times 10^6 \text{Å}^3$

$\Delta S = R\ln(V_g/V_{cr}) = 1.99\text{cal/(g} \cdot \text{mol)} \cdot \ln(0.42 \times 10^6) \approx 25.8\text{cal/(g} \cdot \text{mol)}$

$\Delta H = \Delta G + T\Delta S = 0 + T\Delta S = 300\text{K} \times 25.8\text{cal/(g} \cdot \text{mol)} \approx 7.7\text{kcal/mol}$

习题 5.2

在 27℃、0.1 标准大气压下，气体中的分子浓度如何？

解答：

由于这种气体的单位体积为 $0.25\text{m}^3/\text{mol}$（见习题 5.1 的解答），则在该气体中

的分子浓度等于 $1/(0.25m^3/mol)=1/(250L/mol)=0.004mol/L$。

习题 5.3

在 pH=0 时 H^+离子（确切地说，H_3O^+）之间的特征距离如何？在 pH=7 时呢？

解答：

在 pH=0 时，浓度为 $1mol/L \approx [0.6\times10^{24}$个离子$]/[10^8nm]^3=0.6$ 个离子$/nm^3 \approx$ 1 个离子$/1.7nm^3$

意味着，pH=0 时，H^+离子之间的特征距离为 $(1.7nm^3)^{1/3}\approx1.2nm=12Å$

在 pH=7 时，浓度为 $10^{-7}mol/L\approx1$ 个离子$/(1.7\times10^7nm^3)$，而离子间的特征距离为 $(1.7\times10^7nm^3)^{1/3}\approx257nm=2570Å\approx1/4\mu m$。

习题 5.4

假设某种疏水分子在水和有机溶剂之间的浓度分布 $X_{水}:X_{非极性溶剂}$ 为：0℃时为 $1:(5\times10^4)$，50℃时为 $1:(10\times10^4)$，100℃时为 $1:(4.4\times10^4)$。

试估算该疏水分子从水迁移到有机溶剂中的自由能变化量 ΔG、熵变 ΔS、焓变 ΔH 和比热容变化量 ΔC_p。

解答：

$\Delta G\equiv\Delta G_{水\to非极性溶剂}=-RT\ \ln(X_{非极性溶剂}/X_{水})=RT\ \ln(X_{水}/X_{非极性溶剂})$ [式(5.13)]

0℃（T =273K）时：

$\Delta G=RT\cdot\ln[1/(5\times10^4)]\approx(0.55kcal/mol)\cdot(-10.8)\approx-6.0kcal/mol$

50℃（T =323K）时：

$\Delta G=RT\cdot\ln[1/(10\times10^4)]\approx(0.65kcal/mol)\cdot(-11.5)\approx-7.5kcal/mol$

0℃（T =273K）时：

$\Delta G=RT\cdot\ln[1/(4.4\times10^4)]\approx(0.75kcal/mol)\cdot(-10.7)\approx-8.0kcal/mol$

ΔS 和 ΔH 按下式计算：

$\Delta S=-d(\Delta G)/dT$ [式(5.10)]

$\Delta H=RT^2\cdot[-d(\Delta G/RT)/dT]$ [式(5.11)]

ΔC_p 按下式计算：$\Delta C_p=dH/dT$ [式(5.9)]

ΔS 在区间 0~50℃的中间，即 25℃时为

$-[(-7.5)-(-6.0)]/50°=30cal/(g\cdot mol)$

ΔS 在区间 50~100℃的中间，即 75℃时为

$-[(-8.0)-(-7.5]]/50°=10cal/(g\cdot mol)$

ΔH 在区间 0~50℃的中间，即 25℃（T =298K）时为

$(0.6\times298°)\cdot\{-[(-7.5)/(0.65)-(-6.0)/(0.55)]/50°\}$

$=(0.6\times298°)\cdot(11.5-10.8)/50°\approx2.5kcal/mol$；

ΔH 在区间 50~100℃的中间，即 75℃（T =348K）时为

$(0.7\times348°)\cdot\{-[(-8.0)/(0.75)-(-7.5)/(0.65)]/50°\}$

$=(0.7\times348°)\cdot(10.7-11.5)/50°=-5.5kcal/mol$；

ΔC_p 在区间 25~75℃ 的中间，即 50℃ 时为

[−5.5−(2.5)]/50° = −160cal/(g·mol)

观察图 5-4，我们可以大致从中找到 ΔG、ΔS 和 ΔH 的这些数字，但是符号相反，因为图 5-4 描述的是把分子从有机溶剂迁移到水中，而在本题是讨论的是逆过程：把分子从水中迁移到有机溶剂中。

习题 5.5

（相当复杂，略）

习题 5.6

假设在溶液中有少量的分子 A 和 B，它们可以按照可逆反应 A+B ⟷ AB 进行结合。试证明，当分子 A、B 和 AB 达到平衡时，它们的浓度[A]、[B]和[AB]满足"质量作用定律"，即如下公式：

$$[A]\cdot[B]/[AB]=K_d$$

式中

$$K_d=\exp[(\mu_{AB}^0-\mu_A^0-\mu_B^0)/kT]$$

其中，μ_I^0 为不依赖于分子 I（I = A，B 或 AB）浓度的组分化学式，由分子 I 内部的相互作用以及分子 I 与周围溶剂的相互作用所决定（见习题 17.2）。

解答：

如果已达到平衡，则自由能处于最低点，即它不随着分子 A 和 B 结合成 AB 而变化。如果分子 A、B 和 AB 的浓度很小，则它们的化学势服从公式 $\mu_I=kT\ln[I]+\mu_I^0$，其中 I= A、B 或 AB。意味着，在反应 A+B ⟶ AB 中产生的自由能变化 $\mu_{AB}-\mu_A-\mu_B$ 等于零，即

$$(kT\ln[AB]+\mu_{AB}^0)-(kT\ln[A]+\mu_A^0)-(kT\ln[B]+\mu_B^0)=0$$

和

$$\mu_{AB}^0-\mu_A^0-\mu_B^0=kT\ln\{[A]\cdot[B]/[AB]\}$$

式中，$K_d=\exp[(\mu_{AB}^0-\mu_A^0-\mu_B^0)/kT]$，称为"离解常数"，而 $\mu_{AB}^0-\mu_A^0-\mu_B^0$ 称为粒子 A 和 B 的结合自由能。K_d 值的单位与分子浓度相同（通常为 mol/L）。如果分子主要是处于游离状态（A 和 B），则解离常数很大，如果主要处于结合状态（AB），则离解常数很小。

习题 5.7

如果分子 A、B 和及其结合体 AB 的浓度服从"质量作用定律"，即满足公式 $[A]\cdot[B]/[AB]=K_d$，而且分子 A 的总浓度 [A]+[AB] 为 α，分子 B 的总浓度 [B]+[AB] 为 β。试求平衡浓度[A]、[B]和[AB]。特别是讨论如下情形：解离常数 K_d 很小（与 α 和 β 相比较），而且(1) $\alpha=\beta$；(2) α 远大于 β。

解答：

$$[AB]=(\alpha+\beta+K_d)/2-\{[(\alpha-\beta)/2]^2+K_d(\alpha+\beta)/2+(K_d/2)^2\}^{1/2}$$

$$[A]=\alpha-[AB]$$

[B] $=\beta-$[AB]

附注：由公式$(\alpha-[AB])(\beta-[AB])=K_d[AB]$的平方根（相对于[AB]）很自然地推知，在此情形下所有浓度都是正值。

（1）当 $\alpha=\beta>>K_d$时：

$[AB]\approx\beta-\{K_d\beta\}^{1/2}\approx\beta$;

$[A]=[B]\approx\{K_d\beta\}^{1/2}$.

（2）当 $\alpha>>\beta>>K_d$ 时：

$[AB]\approx\beta\{1-K_d/(\alpha-\beta)\}\approx\beta$;

$[A]\approx\alpha-\beta$

$[B]=K_d[AB]/[A]\approx\beta K_d/(\alpha-\beta)$.

（1）和（2）这两个解在 $\alpha\geqslant\beta>>K_d$ 时可以统一为近似正确的如下形式：

$$[AB]\approx\beta$$

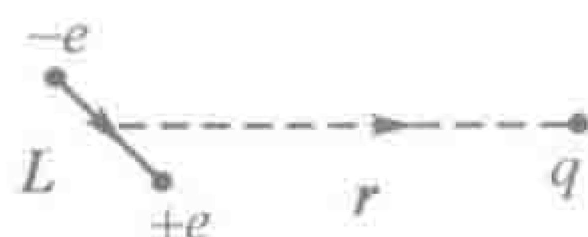
$$[A]\approx\alpha-\beta+\{K_d[AB]\}^{1/2} \tag{5.7.1}$$

$$[B]=K_d[AB]/[A].$$

第6讲

习题 6.1

试求由电荷$+e$和$-e$组成、长度为 L 的电偶极子与电荷 q 之间的静电相互作用自由能。电偶极子与电荷 q 之间的距离 $r\gg L$（如图所示）。媒质的介电系数为 ε。



解答：

$U=\varphi q$，其中 $\varphi=\dfrac{+e}{\varepsilon\,|\boldsymbol{r}-\boldsymbol{L}/2|}+\dfrac{-e}{\varepsilon\,|\boldsymbol{r}+\boldsymbol{L}/2|}$为电偶极子在 q 点处产生的电势。利用展开式

$$\frac{1}{|\boldsymbol{r}-\boldsymbol{L}/2|}=\frac{1}{\sqrt{(\boldsymbol{r}-\boldsymbol{L}/2)^2}}=\frac{1}{\sqrt{r^2-\boldsymbol{rL}+L^2/4}}=\frac{1}{r}\frac{1}{\sqrt{1-\boldsymbol{nL}/r+(L/2r)^2}}\approx$$

$\dfrac{1}{r}\left(1+\dfrac{\boldsymbol{nL}}{2r}\right)$ 和 $\dfrac{1}{|\boldsymbol{r}+\boldsymbol{L}/2|}\approx\dfrac{1}{r}\left(1-\dfrac{\boldsymbol{nL}}{2r}\right)$，其中 $\boldsymbol{n}=\dfrac{\boldsymbol{r}}{r}$ 为从电偶极子中心指向电荷 q 的单位矢量。在展开式中只有对小参数 L/r 的高阶项，则有

$$\varphi=\frac{e}{\varepsilon}\left[\frac{1}{|\boldsymbol{r}-\boldsymbol{L}/2|}-\frac{1}{|\boldsymbol{r}+\boldsymbol{L}/2|}\right]\approx\frac{e}{\varepsilon\cdot r^2}\boldsymbol{nL} \tag{6.1.1}$$

$$U\approx\frac{qe}{\varepsilon\cdot r^2}\boldsymbol{nL}=\frac{q}{\varepsilon\cdot r^2}\boldsymbol{nP} \tag{6.1.2}$$

式中，$\boldsymbol{P}=e\boldsymbol{L}$ 为电偶矩（它从电偶极子的负电荷指向正电荷）。

附注。

(1) 电荷系统的电偶矩通常以 $\boldsymbol{P}=\sum_i e_i\boldsymbol{r}_i$ 的形式进行计算，其中 e_i 和 $\boldsymbol{r}_i$ 是电荷 i 的值(带符号)及其在空间中的坐标。不过这种形式只适合于总电荷 $\sum_i e_i$ 等于 0 的系统；否则系统作为整体移动距离会改变 $\boldsymbol{P}$ (增加 $\boldsymbol{R}\sum_i e_i$)，就会失去物理意义。因此，在一般形式上，更方便的是把电偶矩从电荷系统的"重心" $\boldsymbol{R}_0=\dfrac{\sum_i |e_i|\boldsymbol{r}_i}{\sum_j |e_j|}$ 来进行读数，即把它以 $\boldsymbol{P}=\sum_i e_i(\boldsymbol{r}_i-\boldsymbol{R}_0)$ 进行计算。这样，$\boldsymbol{P}$ 中只含有按一般形式 $\boldsymbol{P}=\sum_i e_i\boldsymbol{r}_i$ 在 $\sum_i e_i=0$ 时进行计算的内容，而如果所有电荷 e_i 的符号都相同，则 $\boldsymbol{P}=\sum_i e_i(\boldsymbol{r}_i-\boldsymbol{R}_0)=0$，这对应于合理的含义。

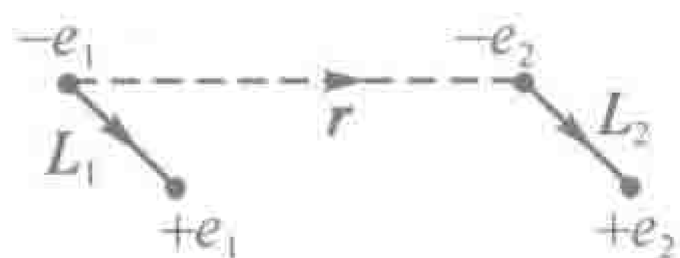
(2) 如果电荷 q 为正(+)，则 $-e$ 被吸引向它，电偶矩 $\boldsymbol{P}$ 努力朝向 $\boldsymbol{n}$ 的反方向，因此 U(以及 $\boldsymbol{nP}$)努力变成负值。如果电荷 q 为负(-)，则 $+e$ 被吸引向它，即电偶矩 $\boldsymbol{P}$ 努力朝向 $\boldsymbol{n}$ 的方向，因此 $\boldsymbol{nP}$ 努力变成正值，而能量 U 仍然为负值。

(3) 至此我们是"从电偶极子的角度"来讨论相互作用。现在"从电荷的角度"来讨论它。因为存在由电荷产生的电场强度，而我们知道 $\boldsymbol{E}_q=q\boldsymbol{n}_{q\to}/\varepsilon r^2$，其中 $\boldsymbol{n}_{q\to}$ 为从电荷 q 朝向场中确定方位的点(本例中为电偶极子所处的位置)的单位矢量，而 $\boldsymbol{n}_{q\to}=-\boldsymbol{n}$ (因为 $\boldsymbol{n}$ 从电偶极子朝向电荷 q)，则同样的能量 U 可以写为如下形式：

$$U=-\boldsymbol{E}_q\boldsymbol{P}\tag{6.1.3}$$

习题 6.2

试求两个电偶极子的静电相互作用的自由能 U：(1) 长为 L_1，由电荷 $+e_1$ 和 $-e_1$ 构成；(2) 长为 L_2，由电荷 $+e_2$ 和 $-e_2$ 构成。这两个电偶极子之间的距离 $r\gg L_1,L_2$ (如图所示)。媒质的介电系数为 ε。



解答：

$U=\dfrac{e_1e_2}{\varepsilon|\boldsymbol{r}|}-\dfrac{e_1e_2}{\varepsilon|\boldsymbol{r}-\boldsymbol{L}_1|}-\dfrac{e_1e_2}{\varepsilon|\boldsymbol{r}+\boldsymbol{L}_2|}+\dfrac{e_1e_2}{\varepsilon|\boldsymbol{r}+(\boldsymbol{L}_2-\boldsymbol{L}_1)|}$。使用在习题 6.1 中相同的展开式，但是现在在它们中含有小参数 L/r 的**两**个高阶项。由于最高阶项约化为

$$\frac{1}{|\boldsymbol{r}+\boldsymbol{L}|}=\frac{1}{r}\frac{1}{\sqrt{1+[2\boldsymbol{n}(\boldsymbol{L}/r)+(L/r)^2]}}$$

$$\approx \frac{1}{r}\left[1+\left(-\frac{1}{2}\right)\left\{2n\frac{L}{r}+\left(\frac{L}{r}\right)^2\right\}+\left[\left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)\frac{1}{2!}\right]\left\{2n\frac{L}{r}+\left(\frac{L}{r}\right)^2\right\}^2+\cdots\right]$$

$$=\frac{1}{r}\left[1-\frac{nL}{r}-\frac{L^2}{2r^2}+\frac{3}{2}\left(\frac{nL}{r}\right)^2\right]$$

式中，$\boldsymbol{n}=\frac{\boldsymbol{r}}{r}$ 为从电偶极子 1 指向电偶极子 2 的单位向量。同样地进行分解并保留项，则

$$U=\frac{e_1e_2}{\varepsilon\cdot r}\left\{1-\left[1+\frac{nL_1}{r}-\frac{L_1^2}{2r^2}+\frac{3}{2}\left(\frac{nL_1}{r}\right)^2\right]-\left[1-\frac{nL_2}{r}-\frac{L_2^2}{2r^2}+\frac{3}{2}\left(\frac{nL_2}{r}\right)^2\right]+\left[1-\frac{n(L_2-L_1)}{r}-\frac{(L_2-L_1)^2}{2r^2}+\frac{3}{2}\left(\frac{n(L_2-L_1)}{r}\right)^2\right]\right\}$$

$$=\frac{e_1e_2}{\varepsilon\cdot r^3}[\boldsymbol{L}_1\boldsymbol{L}_2-3(\boldsymbol{n}\boldsymbol{L}_1)(\boldsymbol{n}\boldsymbol{L}_2)]=\frac{1}{\varepsilon\cdot r^3}[\boldsymbol{P}_1\boldsymbol{P}_2-3(\boldsymbol{n}\boldsymbol{P}_1)(\boldsymbol{n}\boldsymbol{P}_2)] \quad (6.2.1)$$

式中，$\boldsymbol{P}_i=e_i\boldsymbol{L}_i$ 为电偶极子 i 的电偶矩。

以 P_i' 表示电偶矩 $\boldsymbol{P}_i$ 中沿该 $\boldsymbol{n}$ 朝向的分量，以 P_i'' 和 P_i''' 表示垂直于 $\boldsymbol{n}$ 的其他两个分量，则 U 可以写为

$$U=\frac{1}{\varepsilon\cdot r^3}[P_1''P_2''+P_1'''P_2'''-2P_1'P_2'] . \quad (6.2.2)$$

习题 6.3

试求(1)一对彼此远离的“+”和“-”离子(带电荷$+e$ 和$-e$,半径分别为 R^+和 R^-)和(2)一对相互接触的(距离为 $r''=R^++R^-$)“+”和“-”离子的“玻恩能量”，即浸入到水中(从真空中)的自由能。设水的介电系数为 $\varepsilon=80$,忽略 ε 对温度的依赖性。

(3)设“+”和“-”离子以长度为 $r'=(R^++R^-)/2$ 的共价键结合形成电偶极子，求解“玻恩能量”。

在所有情形下都设 $R^+=a$，　$R^-=1.5a$。

解答：

(1)远离的离子：

$$\Delta U^+=\frac{e^2}{2R^+}\left(\frac{1}{\varepsilon}-1\right)=-\frac{e^2}{2a}\left(1-\frac{1}{\varepsilon}\right)=-0.494\frac{e^2}{a};$$

$$\Delta U^-=\frac{e^2}{2R^-}\left(\frac{1}{\varepsilon}-1\right)=-\frac{e^2}{3a}\left(1-\frac{1}{\varepsilon}\right)=-0.329\frac{e^2}{a};$$

$$\Delta U^++\Delta U^-=\frac{e^2}{2R^+}\left(\frac{1}{\varepsilon}-1\right)+\frac{e^2}{2R^-}\left(\frac{1}{\varepsilon}-1\right)=-0.823\frac{e^2}{a};$$

(2) 相距为 r 的两个异号离子：

$$\delta U = \Delta U^+ + \Delta U^- + \left[\frac{-e^2}{r\varepsilon} - \frac{-e^2}{r}\right]$$

$$= e^2\left(-\frac{1}{2R^+} - \frac{1}{2R^-} + \frac{1}{r}\right)\left(1 - \frac{1}{\varepsilon}\right) = 0.9875\,\frac{e^2}{a}\left(\frac{a}{r} - \frac{5}{6}\right);$$

$r = r'' = R^+ + R^- = 2.5a$ 时，$\delta U'' = 0.9875\,\dfrac{e^2}{a}\left(0.4 - \dfrac{5}{6}\right) = -0.423\,\dfrac{e^2}{a}$：相接触的两个离子的浸入自由能实际上等价于单个离子的浸入自由能。

(3) $r = r' = \dfrac{R^+ + R^-}{2}1.25a$ 时，$\delta U'' = 0.9875\,\dfrac{e^2}{a}\left(0.8 - \dfrac{5}{6}\right) = -0.033\,\dfrac{e^2}{a}$：共价键结合的电偶极子的浸入自由能相当低廉。

如果 e 是单个（电子或质子）电荷，而 $a = 1\text{Å}$，则 $e^2/a = 333\text{kcal/mol}$。那么单个离子浸入水中的能量（更确切些：自由能）收益为 100~150kcal/mol（成对键合的离子也相同），而含共价键的电偶极子浸入水中的能量只有几 kcal/mol。

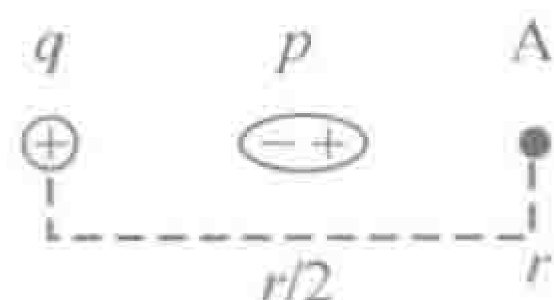
习题 6.4

（相当复杂，略）

习题 6.5

已知，媒质介电系数的升高（与真空中相比）可由沿静电场朝向的电偶极子（该媒质的分子）所产生。也已知，介电系数 $\varepsilon>1$ 会使由电荷所产生的电势（根据公式 $\varphi = \dfrac{q}{\varepsilon r}$，其中 r 为电荷 q 到电势测量点处的距离）减弱。

假设“媒质”是由单个唯一的电偶极子构成，它位于电荷 q 到电场电势测量点 A 的中间（而且该电偶极子自然地沿电荷 q 的电场朝向；如图所示）。



(1) 该电偶极子会使 A 点处的电势比无电偶极子时增大还是减小？

(2) 加强或减弱多少？（为简便起见，假设电偶极子完全沿电荷 q 电场的朝向）

(3) 该完全沿电场朝向的电偶极子在电荷 q 点处所产生的电势如何？

解答：

(1) 有点令人奇怪，也有点违反直觉：这样的电偶极子不是减小，而是**增大**电荷 q 在无电偶极子时在点 A 处产生的电势！

该电偶极子的与电荷 q 相吸引（即与电荷 q 符号相反）的末端，会自然地朝向电荷 q。于是电偶极子中与电荷 q 符号相同的末端会朝向点 A，即它比相反电荷的末端更靠近点 A。这意味着，电偶极子与电荷 q 同号的末端比异号末端对点 A 的作用更**强**，从而**增大**了电荷 q 在点 A 处产生的电势。

(2) 如果电偶极子(长为 L,与电荷 q 同号的电荷 $+e$ 在一端,$-e$ 在另一端)完全地沿电荷 q 的电场朝向,则在原电势 $\varphi=q/r$ 上增加了由电偶极子所生的电势 $\Delta\varphi_A=\frac{+e}{|\boldsymbol{r}/2-\boldsymbol{L}/2|}+\frac{-e}{|\boldsymbol{r}/2+\boldsymbol{L}/2|}=4\frac{P}{r^2}$,其中 $P=|eL|$,为放在电荷 q 和点 A 中点处的电偶极子的电偶矩的大小[式(6.1.1),式(6.1.2)]。

(3) 同一电偶极子在电荷 q 所在点处所产生的电势为 $\Delta\varphi_q=\frac{+e}{|\boldsymbol{r}/2+\boldsymbol{L}/2|}+\frac{-e}{|\boldsymbol{r}/2-\boldsymbol{L}/2|}=-4\frac{P}{r^2}=-\Delta\varphi_A$。结果是电荷 q 的能量会**降低**(如同与极化媒质发生相互作用时一样) $U=q\Delta\varphi_q=-4\frac{qP}{r^2}$。[这一结果是在假设电偶极子热涨落很弱、而且完全沿电荷 q 的电场朝向时得出的;参见式(6.1.1)和式(6.1.2)。]

习题 6.6

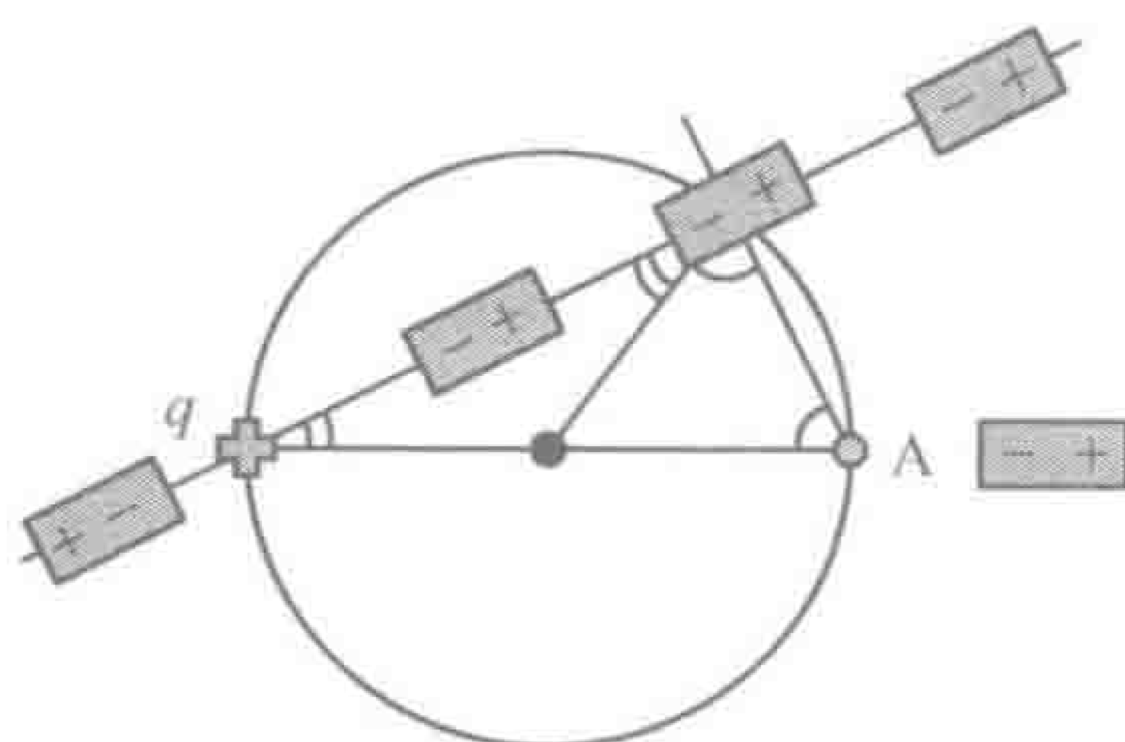
习题 6.5 的解答(1)表明,位于电荷和测量点中间的电偶极子不是减小而是**增大**电荷 q 在无电偶极子时在测量点处产生的电势。这如何与下面的结论:由电偶极子构成的极化媒质会**减小**电荷 q 在无极化媒质时在测量点处产生的电势,相符合?

[提示。这一问题可以表述为:电偶极子放置在何处,能**减小**电荷 q 在无极化媒质时在测量点处产生的电势?]

解答:

为明显起见,假设电荷 q 为正(+)。若电偶极子能使原来由电荷 $+q$ 直接在点 A 处产生的正电势 $\varphi=+q/r$ 减弱,则其"−"端应该比"+"端更**靠近**A 点(而不是如习题 6.5 图中所示的处于 A 和 q 之间的电偶极子,其"+"端朝向 A 点)。

由于电偶极子的朝向沿着电荷 q 所产生的电场,它们所有的"−"端都指向电荷 $+q$(而"+"端朝反方向)。下图显示了通过点 q 和点 A 的任意平面,表明那些位于直径为 q−A 的圆周上的电偶极子,垂直于电偶极子中心与 A 点连接线。(几何学上有这样的定理:基于圆周直径的内切角是直角。)



这意味着,位于以 q−A 线段为半径的球体表面上的电偶极子,如果其"−"和"+"与 A 点的距离相同,则它们不改变 A 点处的电势。

此图显示,位于该球内部、"+"端比"−"端更靠近 A 点的电偶极子(以及显示在习题 6.5 图中所示的电偶极子)会**增大**在 A 点处的原电势 $\varphi=+q/r$。

相反，所有位于该球体**外面**、其“-”端比“+”端更靠近 A 点的电偶极子会减小在 A 点处的原电势 $\varphi=+q/r$（而且这些电偶极子大于该球体内部的电偶极子）。

这意味着，只有那些位于以 q-A 线段为半径的球体外面的电偶极子，会减小原电势 $\varphi=+q/r$。而位于该球体内部的极化媒质完全不会减小，而是增大点 A 处的原电势，这并不矛盾。

习题 6.7

试求电荷在水中的静电相互作用的焓 H 与自由能 G 的关系。电荷相互作用的自由能为 $G\equiv U=qe/\varepsilon r$，而水的介电系数 ε 与温度 T 相关，且 $\varepsilon_{T=273\mathrm{K}}=88$，$\varepsilon_{T=373\mathrm{K}}=55$。

解答：

由于 $G\equiv H-TS$，则 $H=G+TS$。因此，根据式(5.10)，熵 $S=-\mathrm{d}G/\mathrm{d}T$，从而

$$H=G-T\frac{\mathrm{d}G}{\mathrm{d}T}=G\left[1-\frac{T}{G}\frac{\mathrm{d}G}{\mathrm{d}T}\right]=G\left[1-\frac{\mathrm{d}(\ln G)}{\mathrm{d}(\ln T)}\right]$$

由于 $G=\dfrac{qe}{\varepsilon r}$，而且 q、e 和 r 与 T 无关，则 $H=G\left[1+\dfrac{\mathrm{d}(\ln\varepsilon)}{\mathrm{d}(\ln T)}\right]$。这一关系式与媒质无关。与媒质有依赖关系的只是值 $\dfrac{\mathrm{d}(\ln\varepsilon)}{\mathrm{d}(\ln T)}$。尤其是，在 ε 为常数（如在真空中）时，$H=G$。

估算水的 $\dfrac{\mathrm{d}(\ln\varepsilon)}{\mathrm{d}(\ln T)}$：$\dfrac{\ln[\varepsilon_{T=373\mathrm{K}}]-\ln[\varepsilon_{T=273\mathrm{K}}]}{\ln[373\mathrm{K}]-\ln[273\mathrm{K}]}=\dfrac{\ln(55/88)}{\ln(373/273)}=\dfrac{-0.47}{0.31}=-1.5$

这意味着，在水中 $1+\dfrac{d(\ln\varepsilon)}{d(\ln T)}=1-1.5=0.5$，$H=-0.5G$。

于是，**在水中**焓的符号与自由能**相反**，即在**水中**同号电荷靠近时（自由能，即功，当然增大）焓**减小**（而非直觉上看上去的增大）。

习题 6.8

（相当复杂，略）

习题 6.9

（略）

习题 6.10

（相当复杂，略）

第 8 讲

习题 8.1

试求在何温度下，单原子理想气体每个原子的能量等于 ξ。

解答：

$T=\dfrac{2}{3}\dfrac{\xi}{k}$。

计算基于式(8.4) $\frac{1}{T}=\frac{\mathrm{d}S}{\mathrm{d}E}$通过式(8.1) $S=K\ln M(E)$ 。此处 $E=N\xi$,为由 N 个粒子组成的气体的全部能量,而熵由运动部分和坐标部分构成,$S(E)=S^{运动}(E^{动能})+S^{坐标}(E^{势能})$,其中 $E^{动能}+E^{势能}=E$。由于根据定义,理想气体的分子无相互作用,有 $E^{势能}=0$;意味着,$E^{动能}=E$。另一方面,由 N 个粒子构成的单原子理想气体的动能为

$$E^{动能}=\sum_{i=1}^{N}\frac{1}{2}m[\nu_{i,1}^2+\nu_{i,2}^2+\nu_{i,3}^2],$$

式中,m 为分子质量;$\nu_{i,1}$、$\nu_{i,2}$和 $\nu_{i,3}$为粒子 i 沿三维空间坐标运动的速度。意味着,包含有 $3N$ 个坐标矢量($\nu_{1,1},\nu_{1,2},\cdots,\nu_{N,2},\nu_{N,3}$)的长度平方等于 $2E^{动能}/m$,而对由 N 个分子构成的系统在速度空间中可及的点处于半径为 $R=\sqrt{2E^{动能}/m}$ 的 $3N$ 维球体中。该 $3N$ 维球体的面积等于 $C\cdot R^{3N-1}$(系数 C 也指数性地依赖于 N,但我们现在对它不感兴趣),而其相应的熵为 $S^{运动}=k[\ln C+(3N-1)\ln R]$ 。得

$$\frac{1}{T}=\frac{\mathrm{d}S^{运动}}{\mathrm{d}E^{动能}}=k\cdot\frac{\mathrm{d}\left[\ln C+\frac{3N-1}{2}\ln\frac{2E^{动能}}{m}\right]}{\mathrm{d}E^{动能}}=k\cdot\frac{3N-1}{2E^{动能}}=k\cdot\frac{\frac{3N-1}{2N}}{\xi}\tag{8.1.1}$$

合计,在 $N>>1$ 时,有 $T=\frac{2N}{3N-1}\cdot\frac{\xi}{K}=\frac{2\xi}{3K}$。而由于单原子分子三个自由度的每一个所具有的能量 $\xi^{(1)}$平均上等于 $\xi/3$,则 $kT=2\xi^{(1)}$。

补充说明。

这个习题可以用另一种方法来求解:更简短,但需要更多的数学技巧。

从玻尔兹曼分布出发,求得单原子分子每个自由度的平均动能为

$$\xi^{(1)}=\frac{\int_{-\infty}^{+\infty}\frac{m\nu^2}{2}\exp\left(-\frac{m\nu^2/2}{kT}\right)\mathrm{d}\nu}{\int_{-\infty}^{+\infty}\exp\left(-\frac{m\nu^2/2}{kT}\right)\mathrm{d}\nu}=\frac{\int_{-\infty}^{+\infty}\frac{1}{2}kT\cdot\frac{m\nu^2}{kT}\exp\left(-\frac{m\nu^2}{2kT}\right)\mathrm{d}\left(\frac{\sqrt{m}\,\nu}{\sqrt{kT}}\right)}{\int_{-\infty}^{+\infty}\exp\left(-\frac{m\nu^2}{2kT}\right)\mathrm{d}\left(\frac{\sqrt{m}\,\nu}{\sqrt{kT}}\right)}$$
$$=\frac{1}{2}kT\cdot\frac{\int_{-\infty}^{+\infty}x^2\exp\left(-\frac{x^2}{2}\right)\mathrm{d}x}{\int_{-\infty}^{+\infty}\exp\left(-\frac{x^2}{2}\right)\mathrm{d}x}=\frac{1}{2}kT\tag{8.1.2}$$

而由于在三维空间中 $\xi^{(1)}=\frac{1}{3}\xi$,则 $kT=\frac{2}{3}\xi$ 。因此,平均上,粒子的动能为

$$\frac{1}{2}m(\nu_1^2+\nu_2^2+\nu_3^2)=\frac{3}{2}kT\tag{8.1.3}$$

请注意,$\xi^{(1)}=\frac{1}{2}kT$ 中的 $\xi^{(1)}$是每个自由度的平均动能,与粒子质量无关,与其

相互作用的任何假设都无关，甚至与系统的量纲无关，因此它对服从经典力学的任意系统中的所有粒子（条件是这些粒子有很多）都适用。

习题 8.2（有玄机……）

试求由(1)由 N 个不同粒子组成的；(2)由 N 个相同粒子组成的单原子理想气体的坐标熵。每个粒子封闭在体积 V 中。还试求出在这两种情形下粒子的化学势的坐标成分。

解答：

(1) $S=Nk\ \ln V, \mu=-kT\ \ln V$

(2) $S=Nk\ \ln(eV/N), \mu=-kT\ \ln(V/N)$

从第一种情形开始。此时所有粒子各不相同。理想气体的任何粒子的可及坐标空间的体积等于 V（而 V^N 为所有 N 个粒子的构象空间体积）。于是，单个粒子的坐标熵为 $k\ln V$，而 N 个不同粒子的坐标熵为 $S=Nk\ln V$。在恒定体积 V 下添加单个粒子时，系统的自由能的坐标成分增加量为 $\mathrm{d}F = \mathrm{d}E^{势能} - T\cdot k\ln V = -kT\ln V$——由于在理想气体中有 $E^{势能}=0$。于是，化学势的坐标成分 $\mu=-kT\ln V$。（提醒：在当前情形下不能应用式(5.8) $\mu=G/N$，因为它只在系统中包含大量相同粒子时才是正确的。）

如果粒子是相同的（题中第二种情形），则我们需要考虑到，交换任意两个粒子并不改变系统。而由于 N 个粒子可以按 $N! = N\cdot(N-1)\cdot\cdots\cdot 2\cdot 1$ 种方式进行替换（第 1 个粒子可以放到 N 个位置，而第 2 个可以放到 $N-1$ 个位置，依此类推），则非恒同构象空间的体积减少到 $N!$ 分之一，相应地，它的对数减小到 $\ln(N!)\approx N\ln(N/e)$ 分之一（后一变换称为斯特林公式，在 $N\gg 1$ 时成立）。于是，对 N 个相同粒子，有 $S=Nk\ln(eV/N)$。

相应地，该系统的亥姆霍兹自由能的坐标成分为 $F^{坐标} = E^{势能} - T\cdot Nk\ln(eV/N) = -T\cdot Nk\ln(eV/N)$，因而向该系统中添加一个粒子时其自由能增大到 $\mu=\mathrm{d}F^{坐标}/\mathrm{d}N=-kT\ln(V/N)$。

因为在该系统中有大量相同粒子，则 μ 值可以写为如下形式：

$$\mu = \frac{G}{N} = \frac{E - TS + PV}{N} = \frac{-T\cdot Nk\ln(ev/N) + NkT}{N} = -kT\ln\frac{V}{N}$$

（利用了熟知的公式 $PV=NkT$，其中 P 为气体压强，参见下一习题）

在 $N=1$ 时，该化学势自然地与所有粒子不同的情形下仅含一个粒子时相符合。

习题 8.3

试证明，对一定量的理想气体有 PV/T = 常数（其中 V 为体积，P 为压强，T 为温度）。

解答：

根据定义，在恒定的（与外界环境的热交换所保证）温度下物体的自由能变化 $\mathrm{d}F$ 等于施加到物体中的功。同样，根据定义，压强 P 是作用在约束气体表面的单

位面积上的力。在面积 σ 的表面上的作用力为 $P\sigma$。对气体的压缩,即其体积变化 $-\mathrm{d}V<0$,对应于表面 σ 的位移 $-\mathrm{d}V/\sigma$。这一位移需要施加于被压缩物体上的功 $P\sigma\times(-\mathrm{d}V/\sigma)=-P\mathrm{d}V$。于是,$\mathrm{d}F=-P\mathrm{d}V$,且意味着 $(\mathrm{d}F/\mathrm{d}V)_{T不变}=-P$。

自由能分为动能部分和坐标部分,即 $F=(E^{动能}-TS^{运动})+(E^{势能}-TS^{坐标})$。而且 $E^{动能}$ 和 $S^{运动}$ 只与速度有关,而与气体分子的坐标无关;T=常数即外界的温度;对理想气体有 $E^{势能}=0$。只有气体分子在空间的分布熵 $S^{坐标}$ 在气体压缩时发生改变。而根据习题 8.2 的解有 $S^{坐标}=Nk\ln(eV/N)$,并由此得 $\mathrm{d}S^{坐标}/\mathrm{d}V=Nk/V$。于是 $P=-(\mathrm{d}F/\mathrm{d}V)_{T=常数}=T(\mathrm{d}S^{坐标}/\mathrm{d}V)=TNk/V$,且对一定量的理想气体有 $PV/T=Nk=$ 常数。得证。

习题 8.4

(复杂。略)

习题 8.5

服从经典力学的粒子沿一个坐标振动,与该粒子相连的键的弹性能为 $U=1/2Kx^2$,环境温度等于 T,试求该粒子的平均势能 $E^{(1)}$。并求由热涨落而引起的粒子偏离平衡位置的均方根位移 $\sqrt{x^2}$。

解答:

$$E^{(1)}=\frac{kT}{2}\text{ , }\sqrt{x^2}=\sqrt{\frac{kT}{K}}\text{ 。}$$

从玻尔兹曼分布出发,

$$E^{(1)}=+\frac{1}{2}Kx^2=\frac{\int_{-\infty}^{+\infty}\frac{Kx^2}{2}\exp\left(-\frac{Kx^2/2}{kT}\right)\mathrm{d}x}{\int_{-\infty}^{+\infty}\exp\left(-\frac{Kx^2/2}{kT}\right)\mathrm{d}x}=\frac{kT}{2}$$

$$\sqrt{x^2}=\sqrt{\frac{\frac{1}{2}Kx^2}{\frac{1}{2}K}}=\sqrt{\frac{kT}{K}}\qquad(8.5.1)$$

(参见习题解答公式 8.1.2)

请注意,该一维振动的平均势能与一个自由度中计算得的平均动能 $\xi^{(1)}$ 相等。这意味着,一维振动的全部能量(势能+动能)等于 kT。

习题 8.6

(相当复杂。略)

习题 8.7

在某过程的路径中出现高度为 10kcal/mol 的能垒,该过程在 0℃、50℃ 和 100℃ 时变慢多少倍?

解答：

变慢 $\exp[-(10\text{kcal/mol})/RT]$ 倍。即在 0℃ 时变慢到 $\approx 1/(8\times10^7)$，在 50℃ 时变慢到 $\approx 1/(5\times10^6)$，在 100℃ 时变慢到 $\approx 1/(6\times10^5)$。

习题 8.8

根据式(8.20)，粒子在黏度为 η 的液体中的扩散位移 l 的平方与扩散时间 t、kT 值以及粒子直径 D 的依赖关系式为 $L_t^2=t\dfrac{2kT}{\pi Dy}$。

讨论在水中的扩散，27℃ 时 $\eta\approx0.01\text{g/(cm}\cdot\text{s)}$。估算如下粒子从其初始位置移动 1μm（即小细胞的大致尺寸值）的时间：(1) 直径为 $D=3\text{Å}$ 的水分子；(2) 直径 $D=30\text{Å}$ 的小蛋白质分子；(3) 直径 $D=300\text{Å}$ 的大蛋白质复合体。

作为对比，请估算同样的粒子于 27℃ 时在理想气体中飞行 1μm 所需的时间。在所有计算中粒子密度 ρ 取 1.3g/cm^3。

解答：

水分子的扩散时间为 $\sim0.12\times10^{-3}$s，小蛋白质分子为 $\sim1.2\times10^{-3}$s，大蛋白质复合物为 $\sim12\times10^{-3}$s。

无碰撞时飞行 1μm 的时间：水分子为 $\sim1.2\times10^{-9}$s，小蛋白质分子为 $\sim0.4\times10^{-7}$s，大蛋白质复合物为 $\sim1.2\times10^{-6}$s。根据式(8.1.3)，飞行速度 $\nu=(3kT/m)^{1/2}=(18kT/\pi D^3\rho)^{1/2}$。

习题 8.9

假设粒子的"结合位点"是 $\nu_1=30\text{Å}^3$ 的体积，该粒子需要在那里击中自己特定的点 A。还假设粒子在 A 点处于体积 v_1 外面时粒子无法感知位点的靠近程度。

试求在何时间内，直径 $D=30\text{Å}$ 的粒子通过扩散在体积为 $V=1\mu\text{m}^3$ 的水中能找到自己的"结合位点"。27℃ 时，水的黏度为 $\eta\approx0.01\text{g/(cm}\cdot\text{s)}$。

解答：

找到位点的时间：~4s。

由于粒子在 A 点处于体积 ν_1 外面时粒子无法感知位点的靠近程度，则它在落到自己的结合位点之前，需要通过随机扩散来逐个检视所有 $\sim V/\nu_1$ 个体积 ν_1。从一个体积 ν_1 到转换到另一个时需要移动距离 $\sim(\nu_1)^{1/3}$。根据式(8.20)，通过扩散进行的这种基元转换的时间为 $t_1\approx\dfrac{v_1^{2/3}}{2kT/\pi Dy}$。在体积 V 中通过扩散寻找其"结合位点"所花的全部时间为：

$$t\sim t\frac{V}{\nu_1}=\frac{\sqrt[3]{\nu_1^2}}{2kT/\pi D\eta}\frac{V}{\nu_1}=V\frac{\pi\eta}{2kT}\frac{D}{\sqrt[3]{\nu_1}}$$

若 $\eta\approx0.01\text{g/(cm}\cdot\text{s)}$ 且 $T=300\text{K}$，则有 $\pi\eta/2kT=0.4\times10^{-9}\text{s/nm}^3=0.4\text{s}/\mu\text{m}^3$（式 8.21，式 8.22）。

意味着，$t\sim0.4\text{s}/\mu\text{m}^3\times V(D/\nu_1{}^{1/3})$，而且，在 $V=1\mu\text{m}^3$、$D=30\text{Å}$ 和 $\nu_1=30\text{Å}^3$ 时，

$(D/\nu_1^{1/3})\approx 10$,搜索位点的时间约为 4s。

习题 8.10

假设,在质子走近到与质子结合中心有显著的静电相互作用的距离,该处的静电能约为 kT 时,质子能瞬时地与质子结合中心相结合。

27℃时在 pH7 的水中,质子结合中心在多长时间内能捕获质子?

解答:

结合质子的时间为:$\sim 10^{-3}$s。

解释:

在习题 8.9 中已得到公式 $t\sim 0.4\text{s}/\mu\text{m}^3(V\times D/\nu_1^{1/3})$。剩下只需要确定 V、D 和 $\nu_1^{1/3}$。

由于质子 H^+在水中是以$(H_3O)^+$的形式游弋,可以认为,$D\approx 3$Å.

由于 H^+能在距离为 $r\approx 8$Å 处"感知"其最近受体的负电荷(在该处的相互作用能量为 $e^2/\varepsilon r=kT$),则 $\nu_1^{1/3}\approx 16$Å。

$V=1/(10^{-7}\text{mol/L})=10^7\text{L/mol}=10^7\times 10^{15}\mu\text{m}^3/6\times 10^{23}\approx 1/60\mu\text{m}^3$。

第 9 讲

习题 9.1

在一个由 $n=23$ 个链节构成的链段中,在链的螺旋态和线团态之间的平衡常数 K(螺旋/线团) 是多少? 设螺旋延长因子为 $s=1.1$,螺旋初始化因子为 $\sigma=0.001$,而所讨论的链段两端为如下链段:

(1) N 端和 C 端都是线团段;

(2) N 端为线团段,C 端为螺旋段;

(3) N 端为螺旋段,C 端为线团段;

(4) N 端和 C 端都是螺旋段。

对于情形(1)还请估算稳定螺旋:其末端位置在多肽链中不发生涨落的最小长度。

解答:

在情形(1)下,螺旋的形成产生两个(在其始端和末端)螺旋-线团边界,K(螺旋/线团)$=\sigma s^n=0.009$。而且末端位置在多肽链中不发生涨落的稳定螺旋的长度为$-\ln(\sigma)/\ln(s)\approx 73$ 个残基。请注意,末端有涨落的稳定螺旋甚至在 $s=1$ 时也要短得多,仅为(式 9.6)$\sigma^{-1/2}\approx 32$。

在情形(2)和(3)下,螺旋的延长不会改变螺旋-线团边界的数目,K(螺旋/线团)$=s^n=9$。

在情形(4)下,螺旋的形成受限于两个螺旋-线团边界,K(螺旋/线团)$=\sigma^{-1}s^n=9000$。

习题 9.2

(相当复杂,略)

习题 9.3

(相当复杂,略)

习题 9.4

(基于习题 9.2 的延续讨论,略)

习题 9.5

氨基酸残基在链的轮廓长度中的贡献约为 3.5Å,而库恩片段(即链还“记得”自身方向的距离)的长度为 35Å。由 150 个氨基酸残基构成的线团状链的两端之间的距离(即如线团直径)如何?

解答:

线团状链的两端间距离的平方为 $L^2 = 3.5\text{Å} \times 150 \times 35\text{Å} = 18375.5\text{Å}^2$,得:$L \approx 136\text{Å}$.

第 10 讲

习题 10.1

氨基酸残基的平均分子质量为 110Da。

(1) 估算氨基酸残基的平均体积。假设蛋白质的密度等于 1.3g/cm^3。

(2) 由 150 个氨基酸残基构成的球体的直径为多少?

(3) 试求球体直径与由同一链构成的线团半径(在习题 9.5 中已估算)之比。

(4) 假设当分子只占据 1/2 体积时它们开始发生强烈接触,试计算大致在什么浓度下(mmol/mL 或 mol/L)上面所讨论的球体和线团开始发生强烈接触?

解答:

(1) 140Å^3;(2) 34Å;(3) ≈1 : 4;(4) 球体在浓度~650mg/mL (=0.04 摩尔链/L)时开始强烈接触;线团则在~10mg/mL(=0.0006 摩尔链/L)时。

解释:

设 ρ 为蛋白质密度,m 为氨基酸残基的平均分子质量,则其体积为 $V = m/\rho = (110\text{g/mol})/(1.3\text{g/cm}^3) = [110\text{g}/(0.6\times10^{24})]/[1.3\text{g}/(10^{24}\text{Å}^3)] \approx 140\text{Å}^3$。

由 150 个氨基酸残基构成的球体的直径为 $D = (150(140\text{Å}^3\times6/\pi)^{1/3} \approx 34\text{Å}^3$。这是由同一链构成的线团的直径(见习题 9.5)的 1/4,即线团的体积约为球体的 60 倍。

习题 10.2

估算 H^+在 pH7 溶液中的浓度与带电氨基酸残基在与细胞相似,即蛋白质的质量占~10%的溶剂中的浓度(在 mol/L 为单位)之比。

解答:

1 : 2 000 000

如果蛋白质在溶液中的质量占 10%,而带电残基在蛋白质中的比率约为 20%,则这类残基在 1L 溶液中约为 20g,即它们的浓度(分子质量约为 100Da 时)约为 0.2mol/L。而 H^+离子在 pH7 下的浓度为 10^{-7}mol/L。

第 11 讲

习题 11.1

下列序列中，哪个能编码由 α 螺旋编织成的纤维，哪个能编码胶原纤维，哪个能编码 β 结构的纤维蛋白？

(1) -Gly-Ala-Gly-Thr-Gly-Ala-Gly-Thr-Gly-Ala-

(2) -Gly-Ala-Pro-Gly-Pro-Pro-Gly-Thr-Pro-Gly-Ala-Pro-Gly-Pro-Pro-

(3) -Gly-Ala-Glu-Ser-Leu-Gly-Asn-Gly-Ala-Glu-Ser-Leu-Gly-Asn-Gly-Ala-

解答：

(1)：编码 β 结构的纤维蛋白；

(2)：编码胶原纤维；

(3)：由 α 螺旋编织成的纤维。

第 12 讲

习题 12.1

(1) 为什么典型的膜蛋白的外观要么是从膜的一边穿到另一边的 α 螺旋束，要么是也从膜的一边穿到另一边的 β 圆筒？

(2) 处在膜内的可能不是 β 圆筒，而是 β 片吗？

解答：

(a) 因为这些结构能保证在膜内不存在未参与到细胞内相互作用的氢键受体和供体，而在脂质并因而无水的膜环境中，这种“自由的”供体和受体不可能与水结合（像它们在水环境中那样），这种氢键缺陷就会急剧地降低蛋白质结构的稳定性。

(b) 不大可能。因为在 β 片的所有边缘（在不存在 β 圆筒时，它的 β 片边缘彼此粘接）总是有未参与到细胞内相互作用的氢键供体和受体，而这会急剧地降低膜内 β 片的稳定性。

习题 12.2

在膜蛋白内部，在其空间结构的正中间，是否有可能含有大的不规则区段？

解答：

有可能。如果在给定的膜蛋白的内部生成有宽阔的、未被水填充的洞，所讨论的不规则区段就可能摆放在那里。参见图 12-4。

习题 12.3

（困难。略）

习题 12.4

（相当复杂。略）

习题 12. 5

假设，穿过宽 L 的能垒的隧道效应在 0. 001s 内发生。它穿过 2 倍宽相同高度的能垒要花费多长时间？

解答：

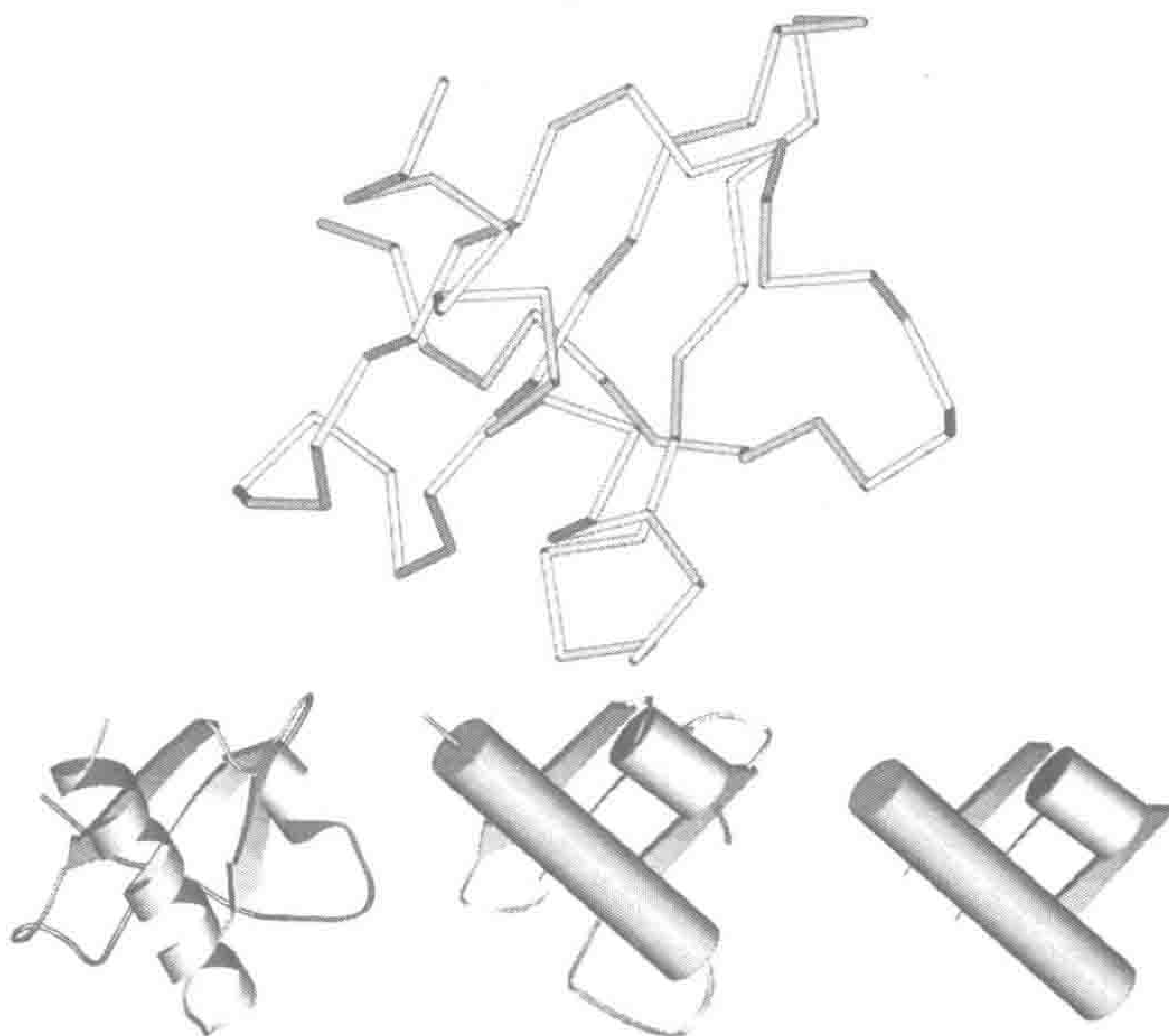
电子穿过给定高度的能垒的隧道效应的概率随该能垒的宽度而指数性降低。如果电子在 0. 001s 内穿过宽为 L 的能垒，则电子在一次电子振动(花费时间 ~10^{-15}s)沿着该能垒开通隧道的概率为 ~10^{-12}。于是电子穿过宽为 $2L$ 的能垒花费的时间为 ~$10^{-15}/(10^{-12}\times 10^{-12})\sim 10^{9}$s。

请注意，以隧道穿过能垒的时间随该能垒的宽度而指数性地升高，而借助热激发的通过时间只正比于能垒的宽度而升高(式 8. 17)。

第 13 讲

习题 13. 1

附图显示了在一个较小的蛋白质中的主链通路。请在其中标出二级结构，并简化为"链折叠模式"，并进一步简化为"结构片段堆码"。



解答：

参见图 13-1、13-3 和立体图 D 型 15(与本题图上下倒过来了)。

习题 13. 2

把如图所示的蛋白质分隔为域。

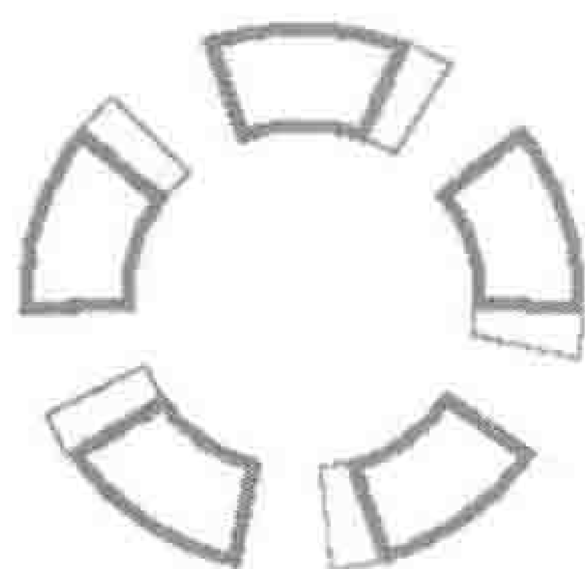


解答：



习题 13.3

请看由 5 个 β 串构成的 β 圆筒(如图所示,阴影表示沿串观看时 β 串通常的向右扭转性)：



(1) 在该 β 圆筒中可能有多少种不同的链折叠模式?

(2) 如果按顺序行进的 β 串彼此反平行(即它们之间的连接线位于圆筒顶端),则在该 β 圆筒中可能有多少种不同的链折叠模式?

(3) 如果在链中按顺序行进的 β 串彼此反平行,而它们之间的连接线不相互交叉,则这样的 β 圆筒有多少种? 画出它们。请注意其中存在有"河曲"和"希腊钥匙"型拓扑的那些。请注意存在由 β 串构成的超螺旋、它在链中的位置及其符号。

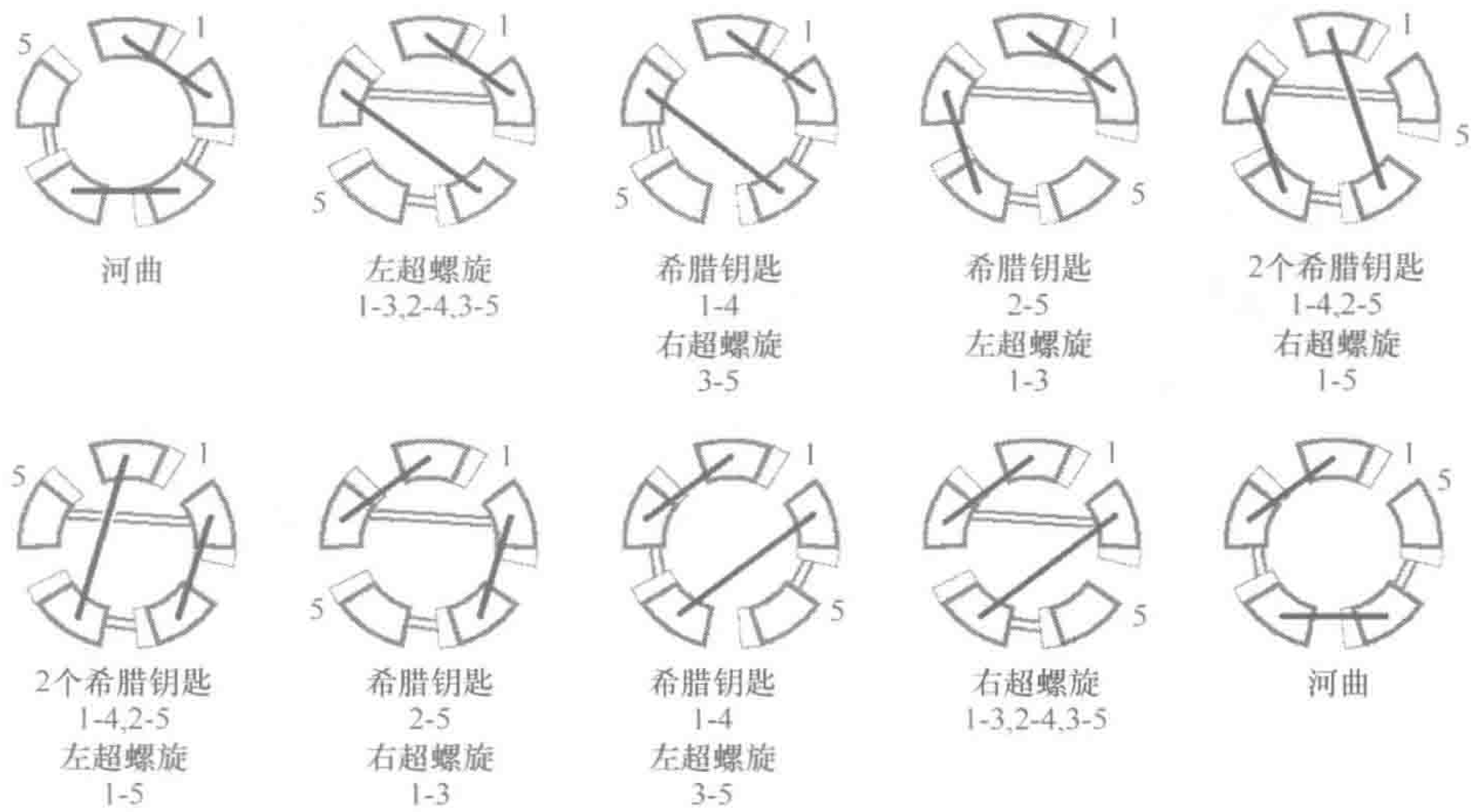
解答：

(1) $4! \times 2^4 = 384$ 种。因为蛋白质链的每个 β 串可以占据圆筒中的 5 个位置

中的任意一个,并可朝向两个方向,但圆筒的第一个(在链中)β 串的位置及其朝向可以任意选取。

(2) 4! =24 种。

(3) 10 种。见下图:



第 14 讲

习题 14.1

在一个域中的二级结构(α 和 β)区段序列看上去如 βαβαβαβαβαβ,在另一个域中看上去如 βββαβββαααα。哪一个属于 α/β 类蛋白质,哪一个属于 α+β 类蛋白质?

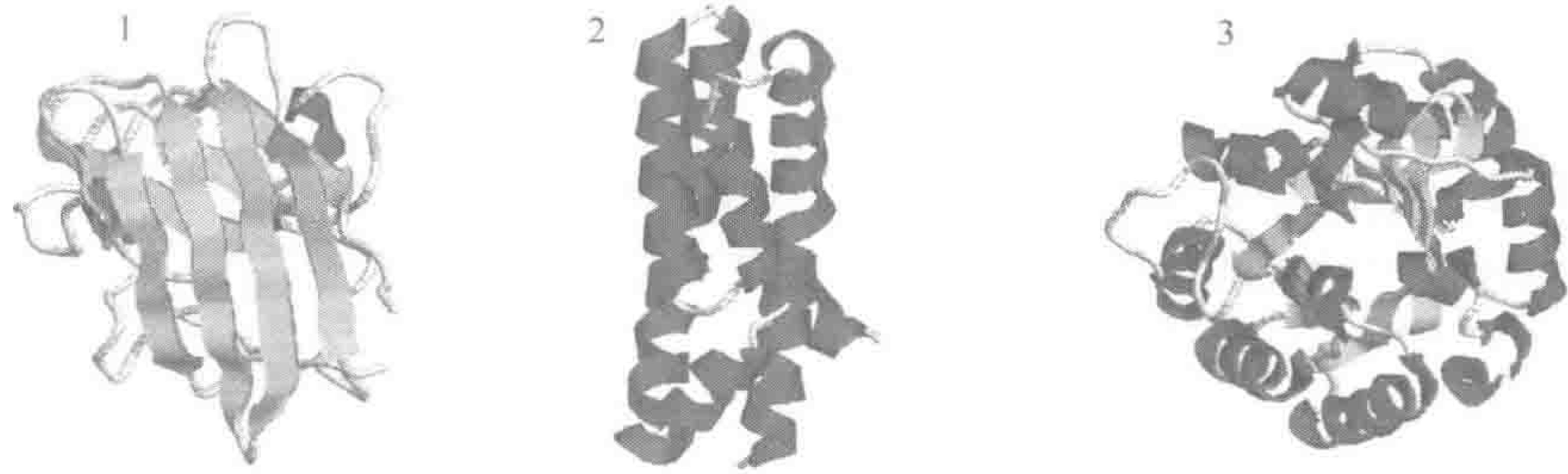
解答:

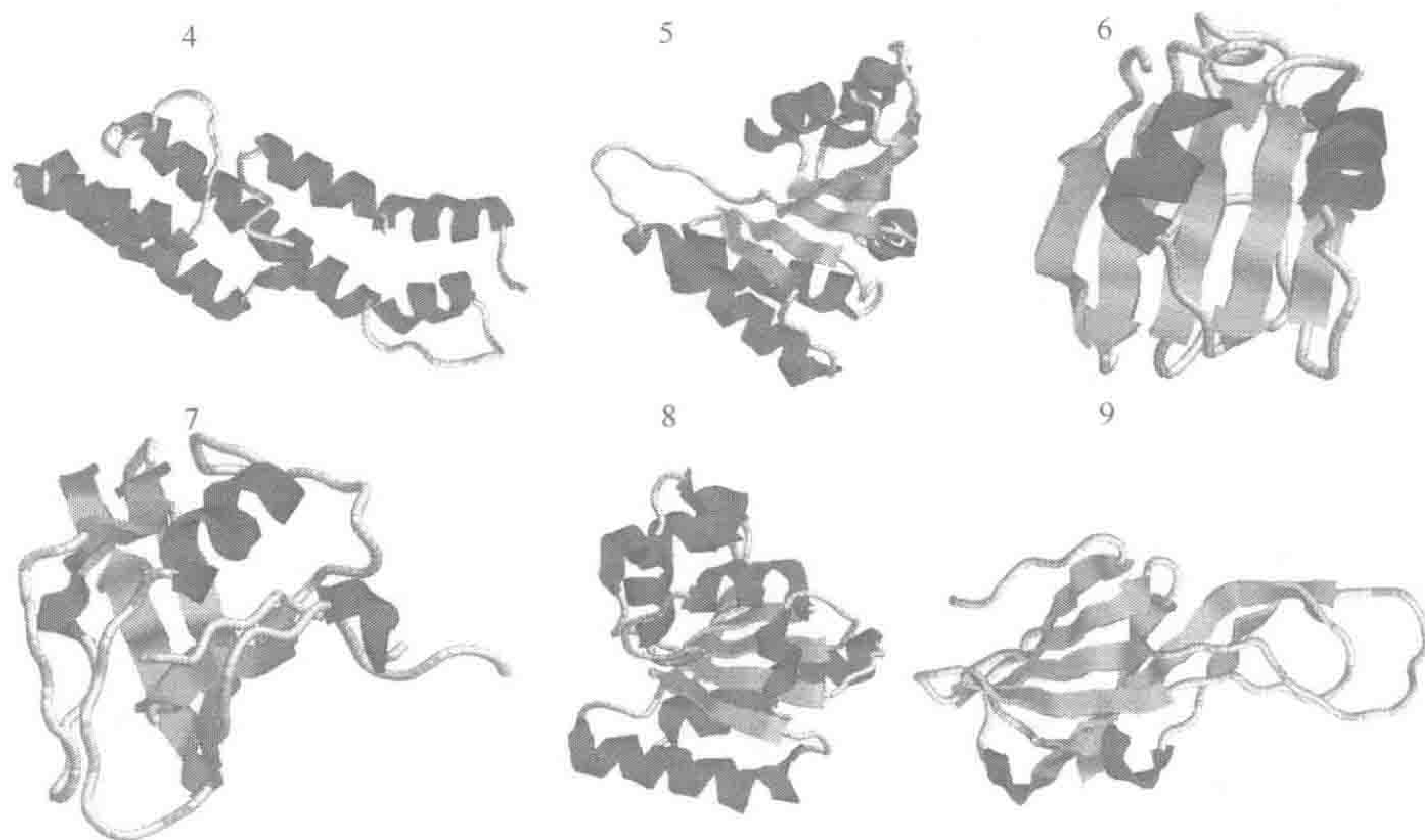
Bαβαβαβαβαβ 属于 α/β 类蛋白质;βββαβββαααα 属于 α+β 类蛋白质。

第 15 讲

习题 15.1

下图中所示的蛋白质域属于水溶性球状蛋白的 4 个主要类别(α、β、α/β、α+β)。





是这样挑选它们的：部分属于该类中最常见的链折叠模式，而部分属于最少见的。

(1) 每一个域分别属于哪一类？

(2) 哪些域具有最常见的链折叠模式，哪些域的链折叠模式是罕见的？

(3) 哪个蛋白质有“Rossmann 折叠”？

解答：

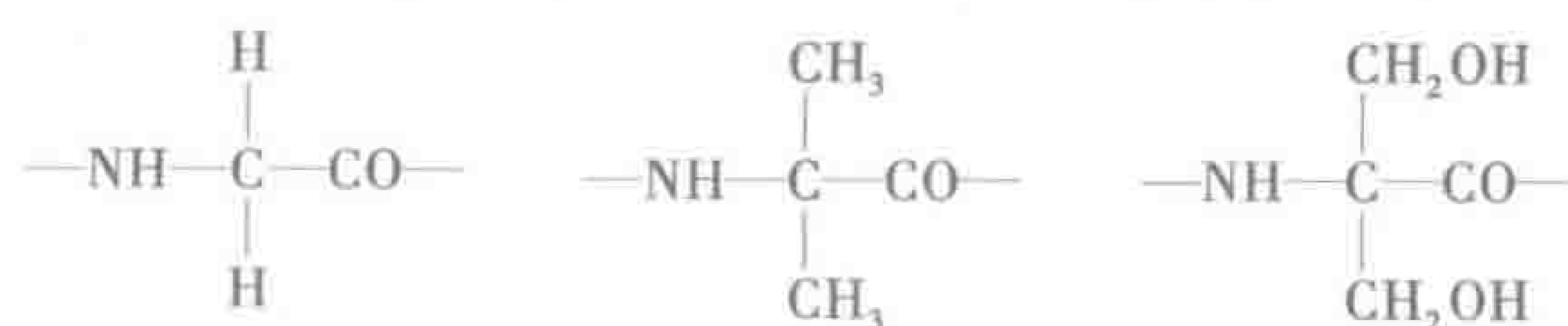
(1) α：2 和 4；β：1 和 9；α/β：3、5 和 8；α+β：6 和 7。

(2) 根据 SCOP 分类法，常见的链折叠模式为 1、2、3、5 和 6；其余的（4、7、8 和 9）为罕见的。通常的“罕见性”原因多少看得出来：在结构 9 中，环线有交叠（在这个并不很紧凑的结构的右边部分）；在结构 4 中，环线深入到 α 层的中心，结构并不很紧凑；在结构 7 中，β 串强烈地扭曲，而且 β 层不是由螺旋覆盖，而是由环线覆盖；而结构 8 看上去完全是“正常的”……

(3) 结构 8 看上去像“Rossmann 折叠”，但在其中没有对 Rossmann 折叠典型的在 β 片中的串次序，并相应地，从 β 串 N 端发出的环线的分布位置（上方的和下方的）有移位。在 Rossmann 折叠中，它们在 β 片（由 6 个 β 串组成）中心发散，而在结构 8 中，仔细观察可以发现，它们靠近与读者接近的 β 片（也是由 6 个 β 串组成）。

习题 15.2

假设，“蛋白质链”由诸如图所示的镜像对称的氨基酸残基构成的。还假设对这种“蛋白质”的天然空间结构的挑选决定于其对给定的一级结构的最大稳定性。



这种“蛋白质”能否具有一个对其稳定性唯一的空间结构？为什么？

解答：

不能。因为如果链是由镜像对称的链节构成的，则链的任何三维折叠都具有不与其重合的镜像相似结构，且这两个结构会具有相同的能量。

这种相似结构的存在不会威胁普通的由 L 型氨基酸构成的蛋白质的空间结构的唯一稳定性，因为普通蛋白质的镜像相似物是由 D 型氨基酸构成的，而 L 型氨基酸转变为 D 型形态要花费极多的时间。

第 16 讲

习题 16.1

(1) 在 $p<1$ 条件下计算 $\sum_{r \geqslant R} p^r$。

(2) 在 $p<1$ 条件下计算 $\sum_{r \geqslant R} r p^r$。

解答：

(1) $\sum_{r \geqslant R} p^r = p^R \cdot \sum_{r \geqslant 0} p^r = \frac{p^r}{1-p}$。因为在 r 趋于无穷大时，级数 p^r 收敛且有 $\sum_{r \geqslant 0} p^r = 1 + \sum_{r \geqslant 1} p^r = 1 + p \cdot \sum_{r \geqslant 0} p^r$，由此得 $(1-p) \sum_{r \geqslant 0} p^r = 1$。

(2) $$\sum_{r \geqslant R} r p^r = p \cdot \frac{\mathrm{d} \sum_{r \geqslant R} p^r}{\mathrm{d}p} = p \cdot \frac{\mathrm{d}\left(\frac{p^R}{1-p}\right)}{\mathrm{d}p} = \frac{p^R}{1-p}\left(R + \frac{p}{1-p}\right)$$

习题 16.2

在漂浮于水中的多肽中，脯氨酸（Pro）经常从顺式（*cis*）转变为反式（*trans*）并反向转变。转变发生在秒级时间范围内，而实验上测定的顺式与反式的比例为1∶10。在蛋白质球体中由顺式到反式（以及由反式到顺式）的转变并没有观察到：顺式脯氨酸占据蛋白质链中的一个位置，而反式脯氨酸占据另一个位置。请估算，在蛋白质球体中应该观察到的顺式脯氨酸与反式脯氨酸的比例如何？

解答：

大约 1∶10（更确切些：蛋白质统计结果表明为 1∶20）。因为天然蛋白质球元件的统计结果与单独取出来的这些元件在 300~400K 的玻尔兹曼统计相似。

习题 16.3

（复杂。略）

第 17 讲

习题 17.1

假设蛋白质分子只有两种态：“0”和“1”。在“0”态，该分子的能量和熵都等于0，在“1”态，该分子的能量为 $E=\mathrm{const}>0$，熵为 $S=\mathrm{const}' \gg k$。

根据范特霍夫判据，求：(1)从"0"态到"1"态转变的特征温度 T^*；(2)该转变的温度范围中的特征宽度；(3)在对应于该转变范围开始点和结束点的温度下，状态"0"和"1" 浓度的特征比例；(4)该转变成为协同转变的条件。

解答：

(1) $T^*=E/S$；

(2) $\Delta T=T^*\cdot(4k/S)$；

(3) 在转变起始点为 $e^2:1\approx 7.4:1$，在结束点为 $1:e^2\approx 1:7.4$；

(4) 条件为 $S\gg 4k$。

解释：

状态"0"和"1"的玻尔兹曼概率为

$$P_0=\frac{1}{1+\exp\left(-\dfrac{E-TS}{kT}\right)}\qquad P_1=\frac{\exp\left(-\dfrac{E-TS}{kT}\right)}{1+\exp\left(-\dfrac{E-TS}{kT}\right)}=1-P_0$$

转变的特征温度 T^* 对应于 $P_1=P_0$，即

$$T^*=E/S$$

$\dfrac{\mathrm{d}P_0}{\mathrm{d}(1/T)}$ 的值表明，P_0 随温度倒数 $(1/T)$ 的变化有多快：

$$\frac{\mathrm{d}P_0}{\mathrm{d}(1/T)}=\frac{\exp\left(-\dfrac{E-TS}{kT}\right)}{\left[1+\exp\left(-\dfrac{E-TS}{kT}\right)\right]^2}\times\frac{\mathrm{d}\left(\dfrac{E-TS}{kT}\right)}{\mathrm{d}\left(\dfrac{1}{T}\right)}=P_0\cdot(1-P_0)\cdot\frac{E}{k}$$

当 $P_0=1/2$ 时，即在温度 T^* 下，这个量取最大值。此时 $\dfrac{\mathrm{d}P_0}{\mathrm{d}(1/T)}=\dfrac{E}{4k}$。在另一方面，$\dfrac{\mathrm{d}P_0}{\mathrm{d}(1/T)}$ 可以写为 $\dfrac{\Delta P_0}{\Delta(1/T)}$ 的形式，其中 ΔP_0 和 $\Delta(1/T)$ 为 P_0 和 $1/T$ 在转变过程中的特征变化。因此，在 $E>0$ 时，$P_0|_{1/T=\infty}\equiv P_0|_{T=0}=1$，而在 $S\gg k$ 时有 $P_0|_{1/T=0}\equiv P_0|_{T=\infty}=\dfrac{1}{1+\exp(S/k)}$，则在转变中 $\Delta P_0=1$，其宽度为

$$\Delta\left(\frac{1}{T}\right)=\frac{1}{E/4k}=\frac{4k}{E},$$

或者，写成更习惯的方式：$\Delta T\approx\dfrac{4k(T^*)^2}{E}=T^*\cdot\dfrac{4k}{S}$。

于是转变范围从 $\dfrac{1}{T'}=\dfrac{1}{T^*}-\dfrac{2k}{E}$ 延伸到 $\dfrac{1}{T''}=\dfrac{1}{T^*}+\dfrac{2k}{E}$。容易求得，在 T' 处有 $P_0=\dfrac{1}{1+\mathrm{e}^2}$ 和 $P_1=\dfrac{\mathrm{e}^2}{1+\mathrm{e}^2}$，在 T''处有 $P_0=\dfrac{\mathrm{e}^2}{1+\mathrm{e}^2}$ 和 $P_1=\dfrac{1}{1+\mathrm{e}^2}$。因此，

在转变开始处：$P_0 : P_1 = e^2 : 1$

在转变结束处：$P_0 : P_1 = 1 : e^2$

转变成为协同转变的条件是它很狭窄，即在 $\Delta\left(\frac{1}{T}\right) \ll \frac{1}{T^*}$ 条件下（或 $\Delta T \ll T^*$），这发生在 $S \gg 4k$ 的条件下。

习题 17.2

试提出用于描述粒子在浓度很低的溶液中的化学势的一般公式。

解答：

$\mu = kT\ln[C] + \mu^0$，其中，C 为所研究的粒子在溶液中的浓度，与此浓度无关的化学势 μ^0 取决于温度和单个粒子与其周围溶剂之间的相互作用。

解释。在习题 8.2 中求得了粒子在气体中的化学势的坐标成分的公式：$\mu^{坐标} = -kT\ln(V/N)$，其中，T 为温度，N 为包含在体积 V 中的粒子数。根据定义，值 $C = N/V$ 为粒子密度。

除了坐标成分之外，气体中的粒子的化学势还包含与浓度 C 无关的运动成分和分子内相互作用的自由能，而对于液体中的粒子，还额外有它与其周围溶剂的相互作用的自由能。这些项都包含在 μ^0 中。

习题 17.3

假设蛋白质分子只有"0"和"1"两种状态。在"0"态其自由能等于 0，且漂浮在溶液中的粒子 A（如质子）未与它结合，在"1"态它结合有 n_A 个粒子 A，因而该分子的自由能变成等于 $F = F_1 - n_A\mu_A > 0$（其中 F_1 为蛋白质在状态"1"并与 n_A 个粒子 A 相结合时的私有自由能，而 $-n_A\mu_A$ 项是考虑到 n_A 个粒子 A 被从溶液中提取出来，它们在溶液中的化学势为 μ_A）。

转变是由值 μ_A 的变化引起的：它在 $\mu_A^* - \frac{\Delta\mu_A}{2} \to \mu_A^* + \frac{\Delta\mu_A}{2}$ 的过程中发生的，其中 $\mu_A{}^*$ 对应着转变中点。

试求对这种过程的范特霍夫相似公式。

解答：

$n_A\Delta\mu_A = 4kT$，其中 T 为环境温度；$\mu_A{}^* = F_1/n_A$。

解答过程与习题 17.1 相似。状态"0"和"1"的玻尔兹曼概率为 $P_0 = \frac{1}{1 + \exp[-(F_1 - n_A\mu_A)/(kT)]}$，$P_1 = 1 - P_0$。值 $\frac{dP_0}{d\mu_A} = -P_0 \cdot (1 - P_0) \cdot \frac{n_A}{kT}$ 表示 P_0 随粒子 A 的化学势而变化有多快，它在 $P_0 = 1/2$ 时取最大值。此时 $F_1 = n_A\mu_A{}^*$，而 $\frac{dP_0}{d\mu_A} = -\frac{n_A}{4kT}$。另一方面，$\frac{dP_0}{d\mu_A} = \frac{\Delta P_0}{\Delta\mu_A}$，其中 ΔP_0 和 $\Delta\mu_A$ 为 P_0 和 μ_A 在转变过程中的特征变化量。因此在系统中完全不存在粒子 A（此时 $C = 0$，因此根据习题 17.2，有 $\mu_A = -\infty$）时 $P_0 = 1$，而在 $C \to \infty$（此时 $\mu_A \to \infty$）时有 $P_0 \to 0$，则在从 $C = 0$ 到 $C = \infty$

的转变过程中有 $\Delta P_0=-1$，而且

$$n_A\Delta\mu_A=4kT\ 。$$

因此，只要知道了与转变相对应的粒子 A 化学势的变化(例如，对于 H^+ 有 $\Delta\mu=2.3kT\cdot\Delta pH$)，我们可以求出与协同转变单位相结合的粒子 A 的数量 $n_A=4kT/\Delta\mu_A$，进而，如果我们知道多少个粒子 A 被蛋白质分子整体地吸引，则可以知道这个协同转变单位是否与蛋白质分子整体相吻合。比较在存在蛋白质和不存在蛋白质时使化学势 μ_A 升高 $\Delta\mu_A$ 所需的粒子 A 数量，并把这个差值除以溶液中的蛋白质分子数目，则可以知道蛋白质所吸收的粒子 A 的数量。在“粒子 A”为质子的情形下，这种方法被称为化学势滴定测量法。

习题 17.4

蛋白质在温度 $T^*=350$ K 时熔化(以“全或无”型)，而且熔解范围宽度 $\Delta T=7°$。

(1) 蛋白质在熔点处的天然态与变性态的焓差值 ΔH 为多少？蛋白质在熔点处的天然态与变性态的熵差值 ΔS 为多少？

(2) 蛋白质在熔化温度 T^* 处的天然态与变性态的自由能差值 ΔG 为多少？

(3) 蛋白质在比熔化温度 T^* 低 10° 处的天然态与变性态的自由能差值 ΔG 为多少？

解答：

(1) 根据范特霍夫关系式，$\Delta T=4k(T^*)^2/\Delta E$，其中，$\Delta E$ 为协同单位在“全或无”型转变中的焓升高量。意味着，蛋白质变性态的焓高于天然态的焓，而天然态与变性态的焓差值 $\Delta H=-\Delta E=-4kT^*\times(T^*/\Delta T)=-4\times(0.7\text{kcal/mol})\times50=-140\text{kcal/mol}$。而由于 $\Delta S=\Delta H/T^*$，则 $\Delta S=(-140\text{kcal/mol})/(350\ \text{K})=-400\text{cal/(mol}\cdot\text{K)}$。

(2) 0，根据熔点的定义！

(3) 由于 $\Delta S=-\dfrac{\partial(\Delta G)}{\partial T^*}$，则 $\Delta G(T^*-10°)\approx\Delta G(T^*)-10°\times(-\Delta S)=0-10°\times400\text{cal/(mol}\cdot\text{K)}=-4\text{kcal/mol}$。

习题 17.5

(并不复杂，但需要计算。略)

习题 17.6

(复杂。略)

习题 17.7

(不难，但需要计算。略)

第 18 讲

习题 18.1

单个蛋白质分子的变性实验是这样做的：分子的一端被化学性地缝合到底板上，而另一端缝合到原子力显微镜的探针上，对该探针施加某种使蛋白质解折叠的

作用力。

请估算使由 100 个氨基酸残基构成的蛋白质解折叠所需要的作用力。与线团相比,该蛋白质的天然态的稳定性为 10kcal/mol。假设观测时间无限制。

解答:

为了使该蛋白质解折叠,需要消耗功 $A = 10\text{kcal/mol}$。这个功需要花费在等于拉直链与蛋白质球体的线性大小差值的路程 L 上,即 $L \approx 350\text{Å} - 40\text{Å} \approx 300\text{Å}$。所求的作用力 $F = A/L = (10\text{kcal/mol})/300\text{Å} \approx (42\times10^3\text{J}/6\times10^{23})/(300\times10^{-10}\text{m}) \approx 2\times10^{-12}\text{N} = 2\text{pN}$。

习题 18.2

多少种链节才能使链单值地确定其空间结构的问题,可以变形为如下习题的形式:

假设,由 N 个链节构成的链具有 M 种折叠形式,有机会成为对某种链节序列的最优结构的其他性质都相同(还假设这些折叠形式的基本性质如紧致度、二级结构的存在等等都相同)。

链节种类的数量 K 要有多少,才能保证每个链节序列只存在一个具有对该序列的能量最小的折叠形式?

解答:

$K \geqslant M^{1/N}$

如果有 K 种链节,则由 N 个链节构成的链具有 K^N 种可能的序列。如果 $K^N < M$,且每个序列只有一个最优折叠形式,则这 K^N 个最优折叠形式无论如何都不能覆盖所有 M 个可能的折叠形式(假设从只要对某种序列有机会能成为最优折叠形式的角度来说,它们是等价的)。这一矛盾表明,K^N 应该超过 M,即 $K \geqslant M^{1/N}$。

习题 18.3

(复杂。略)

第 19 讲

习题 19.1

假设从 2^{100} 个结构中挑选一个以 3 个动力学步骤进行。在第一步,链从由 2^{70} 个结构组成的对应于一级中间体的子集中挑选出一些结构。第二步,不超出这个子集,链从由 2^{40} 个结构组成的对应于二级中间体的子集中挑选出一些结构。在最后一步,挑选出唯一的结构。

假设,在每一个步骤的框架内,结构挑选是通过完整的轮换来进行的,而且不存在“伪”能量最小值(不与“正确的”中间体相应的“正确”结构的折叠)。

(1) 设从一个结构转变到另一个结构占用时间 10^{-10}s,请估算整个过程需占用多长时间。

(2) 在第一步、第二步和第三步需要能量降低多少,以便链不会返回到前一步?(假设 $kT = 0.6\text{kcal/mol}$)

解答:

(1) 在第一步中的挑选需要从 $2^{100}/2^{70}=2^{30}\approx10^{9}$ 个结构中挑选一个;这对找到一级中间体是足够了。如果链不会“陷入”到不对应于“正确的”中间体的能量最小值(否则会花费更多时间),则这种过程花费 $10^{-10}\text{s}\times10^{9}\sim0.1\text{s}$。在第二步中的轮换(找到二级中间体)需要从 $2^{70}/2^{40}=2^{30}\approx10^{9}$ 个结构中挑选一个;在相同的条件下,它也花费 ~0.1s。在第三步的轮换中需要从 $2^{40}\approx10^{12}$ 个结构中挑选出一个:它花费 $10^{-10}\text{s}\times10^{12}\sim100\text{s}$(在不存在“错误的”能量最小值的情形下)。总计:最小大约 100s。

(2) 在第一步中:至少要比 $kT\ln(2^{30})\approx12.5\text{kcal/mol}$ 要大 kT。第二步中需要一样多。在第三步中:至少比 $kT\ln(2^{40})\approx16.7\text{kcal/mol}$ 要大 kT。总计:不少于 $3kT+kT\ln(2^{30})\approx44\text{kcal/mol}$。

习题 19.2

(复杂。略)

习题 19.3

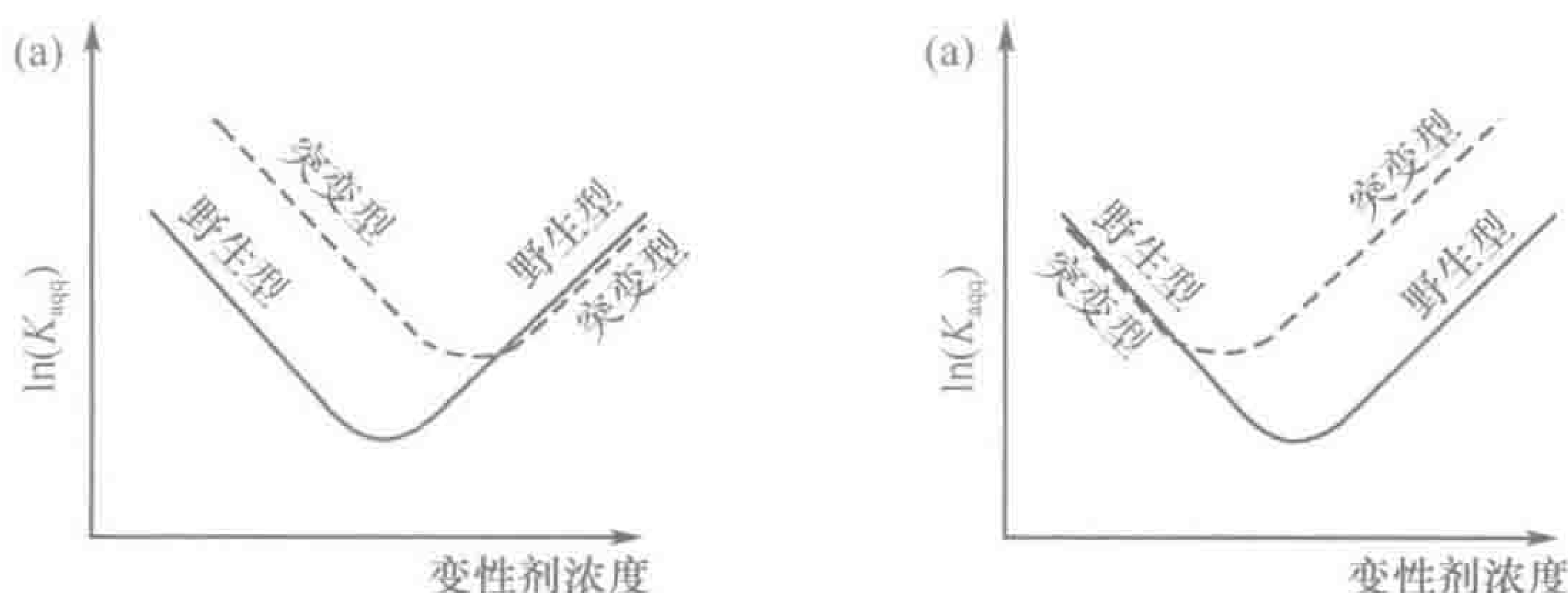
(略)

第 20 讲

习题 20.1

假设,在某些蛋白质中一个氨基酸残基的突变会产生 V 形图相对于“野生型”蛋白质 V 形图如下图(a)中的偏移,而另一个残基的突变会产生如下图(b)中的偏移。

对这些残基对蛋白质折叠核的参与程度,我们能说什么?



解答:

图(a)中所示的 V 形图偏移说明了在蛋白质解折叠速度(k_u)几乎无变化时折叠速度(k_f)的变化。这意味着,突变既影响了折叠核的稳定性,也影响了蛋白质天然结构的稳定性,这由 k_f/k_u 比例关系测定。这意味着,第一个所讨论的残基完全地参与到折叠核中。

图(b)所示的相反情况说明第二个所讨论的残基完全不参与折叠核。

习题 20.2

在溶菌酶天然态和解折叠态的热力学平衡点,每个溶菌酶分子都经历“全或

无”型转变，蛋白质溶液接近平衡状态的速率大约为 $0.003s^{-1}$（图 20-4）。

（1）在该速度下蛋白质折叠速率是多少？

（2）蛋白质折叠的特征时间是多少？

（3）蛋白质解折叠的速度和特征时间是多少？

（4）如果蛋白质最初处于 8mol/L 的尿素溶液中，在时刻“0”处该溶液被水稀释到尿素浓度为 4.2mol/L，该浓度对应着蛋白质天然态和解折叠态的热力学平衡点，则蛋白质溶液接近平衡点的过程的动力学是什么样？

（5）如果在上述实验中，我们长时间地注视单个蛋白质分子，则该分子的折叠-解折叠过程是什么样？

（6）如果蛋白质最初处于纯水中，在时刻“0”处向该溶液中加入尿素，直到尿素浓度为 4.2mol/L，该浓度对应着蛋白质天然态和解折叠态的热力学平衡点，则蛋白质溶液接近平衡点的过程的动力学是什么样？

（7）如果在上述实验中，我们长时间地注视单个蛋白质分子，则该蛋白质的单个分子的折叠-解折叠过程是什么样？

解答：

（1）蛋白质折叠速率为：1/2×（接近平衡的速率）= $0.000\ 15s^{-1}$。

（2）蛋白质折叠的特征时间为：1/（蛋白质折叠速率）= $6.7\times10^3s\approx2h$。

（3）在热力学平衡点，解折叠速率等于折叠速率，即等于 $0.000\ 15s^{-1}$，解折叠特征时间为 $1/0.000\ 15s^{-1}=6.7\times10^3s\approx2h$。

（4）在尿素稀释的“0”时刻，天然蛋白质的比率为 0，然后（在天然蛋白质和变性蛋白质的热力学平衡点）该比率为 $0.5\times[1-\exp(-0.000\ 15\cdot t)]$，其中 t 为从“0”时刻之后经过的时间（以秒为单位）。

（5）在尿素稀释之后，单个蛋白质分子保持为解折叠态约 2h，然后几乎瞬间折叠，再保持为折叠态约 2h，再几乎瞬间解折叠，再保持为解折叠态约 2h，再几乎瞬间折叠……如此继续。（附注：“约 2h”表示，对某些分子该时间为 0.5h，对另一些为 1h，对另一些为 2h，对另一些为 4h，平均上为 2h。）

（6）在添加尿素的“0”时刻，天然蛋白质的比率为 1，然后为 $0.5\times[\exp(-0.000\ 15\cdot t)+1]$，其中，$t$ 为从“0”时刻之后经过的时间（以秒为单位）。

（7）在添加尿素之后，单个蛋白质分子保持为折叠态约 2h，然后几乎瞬间解折叠，再保持为解折叠态约 2h，再几乎瞬间折叠……如此继续。

习题 20.3

假设在室温下蛋白质发生从解折叠态到天然态的“全或无”型转变，而且其折叠速率等于 $3s^{-1}$，而天然态比解折叠态更稳定 7.5kcal/mol。

（1）该蛋白质（折叠之后）保持为折叠态的特征时间是多少？

（2）该蛋白质（解折叠之后）保持为解折叠态的特征时间是多少？

（3）如果我们长时间观察单个蛋白质分子，在给定条件下的折叠-解折叠过程

是什么样?

解答:

(1) 大约 1 昼夜。如果在室温下天然态比解折叠态更稳定 7.5kcal/mol,则蛋白质的解折叠速率比折叠速率要慢,为 $\exp[(7.5\text{kcal/mol})/(0.6\text{kcal/mol})]=1/(2.5\times10^5)$。意味着,蛋白质解折叠速率为 $(3\text{s}^{-1})/(2.5\times10^5)\approx1/(83\ 000\text{s})\approx1$ (昼夜)$^{-1}$,而其特征时间为 1/[1(昼夜)$^{-1}$]=1 昼夜。

(2) 约为 $1/(3\text{s}^{-1})=0.33\text{s}$。

(3) ……保持为解折叠态大约 0.3 秒 → 几乎瞬间折叠 → 保持为折叠状态约 1 昼夜 → 几乎瞬间解折叠 → 保持为解折叠态约 0.3 秒 → 几乎瞬间折叠 → ……

第 21 讲

习题 21.1

对于由 $N=100$ 个氨基酸残基构成的蛋白质,请在如下条件下,非常近似、只准确到数量级地估算在变性点区域附近的折叠时间:

(1) 我们完全不知道该蛋白质的空间结构;

(2) 我们已知在其 α 螺旋中有 70 个氢键;

(3) 已知在其 α 螺旋中有 70 个氢键,在室温下其天然结构的稳定性为 6kcal/mol,而且已知蛋白质折叠时没有可见的折叠中间体。

解答:

(1) 平均上,$\sim10\text{ns}\times\exp(100^{2/3})\sim10\text{s}$;定性上的范围从 $\sim10\text{ns}\times\exp(0.5\times100^{2/3})\sim10^{-3}\text{s}$ 到 $\sim10\text{ns}\times\exp(1.5\times100^{2/3})\sim10^5\text{s}$。

(2) $\sim10\text{ns}\times\exp(30^{2/3})\sim0.16\text{ms}$。

(3) 粗略地说,由于折叠核包含蛋白质的一半,则核的稳定性大约高 3kcal/mol,因此:$\sim0.16\text{ms}\times\exp[-(3\text{kcal/mol})/(0.6\text{kcal/mol})]\sim1\mu\text{s}$。

习题 21.2

在溶液中生长的晶体的自由能可表示如下:

$$F=Mf+6M^{2/3}\psi, \tag{21.2.1}$$

式中,M 为晶体中的分子数;f 为晶体相分子相对于液相分子的自由能;ψ 为晶体相分子在其表面时的附加自由能。

仅为简便起见,设分子从液相嵌入到新生晶体的特征时间 τ 既与晶体尺寸无关,也与晶体表面形状无关。

(1) 估算在给定空间点上晶体相的特征初始化时间 t_{init}。

(2) 求 t_{init} 与稳定晶体最小尺寸 M_{st} 之间的联系。

(3) 估算微米尺寸,即由 $N_1\sim10^{10}$ 个分子构成的液滴中产生第一个晶体的特征时间 t_1。

(4) 在室温下进行数值估算:① $f=-0.2\text{kcal/mol}$, $\psi=0.3\text{kcal/mol}$, $\tau=10^{-11}\text{s}$;

② $f=-0.1\text{kcal/mol}, \psi=0.3\text{kcal/mol}, \tau=10^{-11}\text{s}$。

解答：

由 $Mf+6M^{2/3}\psi=0$ 推导得稳定晶体的最小尺寸为 $M_{st}=\left(\frac{6\psi}{-f}\right)^3$

由 $\frac{d(Mf+6M^{2/3}\psi)}{dM}=0$ 推导得晶体的晶核尺寸为 $M_1=\left(\frac{4\psi}{-f}\right)^3=\frac{8}{27}M_{st}$，其相对于液体的自由能为 $\Delta F^{\#}=2\psi(M_1)^{2/3}=\frac{8}{9}\psi(M_{st})^{2/3}$。

在给定空间点上晶体相的特征初始化时间为

$$t_{init}=\tau\cdot\exp\left(\frac{+\Delta F^{\#}}{kT}\right)=\tau\cdot\exp\left(\frac{8\psi}{9kT}M_{st}^{2/3}\right)$$

$$t_1\sim\frac{t_{init}}{N_1}$$

在 $T=300\text{K}$（即 $kT=0.6\text{kcal/mol}$），$\psi=0.3\text{kcal/mol}$，$\tau=10^{-11}\text{s}$ 时进行数值估算：

若 $f=-0.2\text{kcal/mol}$，则 $M_{st}=9^3=729$，$t_{init}=10^{-11}\text{s}(\exp[4/9(9^2]\sim10^5\text{s}$，$t_1\sim10^{-5}\text{s}$。

若 $f=-0.1\text{kcal/mol}$，则 $M_{st}=18^3=5832$，$t_{init}=10^{-11}\text{s}(\exp[4/9(18^2]\sim10^{52}\text{s}$，$t_1\sim10^{42}\text{s}$。

最后一个估算表明，只有在较浓的液体中或在其表面含有某种不均匀性或表面含有能易化晶体初始化的杂质时，才能在液相和晶体相的平衡点附近处实际观察到结晶化。

第 22 讲

习题 22.1

把序列 B 和 C 相对于序列 A 进行比对。

A： P D C A P T E P D F V S D T

B： P D S C T E P S A F V T D T

C： P V A C A V E P S V F V S S D

设在比对中，每个氨基酸相同得分+1，不同得分 0，而任意氨基酸残基空位得分-1(除了末端的空位：不得分)。

解答：

A 与 B 的比对为

```
  A:   P  D  –  C  A  P  T  E  P  –  D  F  V  S  D  T
  B:   P  D  S  C  –  –  T  E  P  S  A  F  V  T  D  T
得分:  +1+1–1+1–1–1+1+1+1–1  0+1+1  0+1+1  =  +6
```

A 与 C 的比对为

```
  A:   P  D  –  C  A  P  T  E  P  –  D  F  V  –  S  D  T
  C:   P  V  A  C  A  V  –  E  P  S  V  F  V  S  S  D  –
得分:  +1  0–1+1+1  0–1+1+1–1  0+1+1–1+1+1  0  =  5
```

或

A:	P	–	D	C	A	P	T	E	P	D	–	F	V	S	–	D	T
C:	P	V	A	C	A	–	V	E	P	S	V	F	V	S	S	D	–

得分：+1−1 0+1+1+1−1 0+1+1 0−1+1+1+1−1−1 0=5

或……(如此等等)

(在给定参数下非唯一的比对加以下划线显示；氨基酸的重合以黑体字标出)。

习题 22.2

在上一题解答的基础上：

(1) 对序列 A、B 和 C 进行多序列比对；

(2) 寻找序列 A、B 和 C 中的保守序列；

(3) 绘出序列 A、B 和 C 的一级结构谱(profile)

(4) 使用该一级结构谱，识别下面两个序列(1 和 2)哪一个与序列 A、B 和 C 同源。

1： F V S F A V A A A S V F S S A T

2： P F F C E E E P T T T F V A D A

比对参数与上一题相同：每个氨基酸相同得分+1，不同得分 0，而任意氨基酸残基在序列或一级结构谱位置中的空位得分−1；附加条件：氨基酸残基在序列和一级结构谱中的缺位相同得分+0.5。假设比对得分超过 3.5 时能证明同源性。

解答：

(1) 比对有多种。它往往允许取非单一的成对比对(有非单一比对的位置加下划线显示)

A:	P	D	–	C	A	P	T	E	P	–	D	F	V	–	S	D	T
B:	P	D	S	C	–	–	T	E	P	S	A	F	V	–	T	D	T
C:	P	V	A	C	A	–	V	E	P	S	V	F	V	S	S	D	–

(2) 共识(consensus)：

		–								V						
	V	S		–	–	V			–	A			–	T		–
P	D	A	C	A	P	T	E	P	S	D	F	V	S	S	D	T

(3) 一级结构谱：

氨基酸残基：			–								V						
残基相同的得分：			$+\frac{1}{3}$								$+\frac{1}{3}$						
氨基酸残基：		V	S		–	–	V			–	A			–	T		–
残基相同的得分：		$+\frac{1}{3}$	$+\frac{1}{3}$		$+\frac{1}{3}$	$+\frac{2}{3}$	$+\frac{1}{3}$			$+\frac{1}{3}$	$+\frac{1}{3}$			$+\frac{2}{3}$	$+\frac{1}{3}$		$+\frac{1}{3}$
氨基酸残基：	P	D	A	C	A	P	T	E	P	S	D	F	V	S	S	D	T
残基相同的得分：	+1	$+\frac{2}{3}$	$+\frac{1}{3}$	+1	$+\frac{2}{3}$	$+\frac{1}{3}$	$+\frac{2}{3}$	+1	+1	$+\frac{2}{3}$	$+\frac{1}{3}$	+1	+1	$+\frac{1}{3}$	$+\frac{2}{3}$	+1	$+\frac{2}{3}$

或以另一种记录方式：

A:	0	0	$\frac{1}{3}$	0	$\frac{2}{3}$	0	0	0	0	0	$\frac{1}{3}$	0	0	0	0	0	0
C:	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
D:	0	$\frac{2}{3}$	0	0	0	0	0	0	0	0	$\frac{1}{3}$	0	0	0	0	1	0
E:	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
F:	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
P:	0	0	0	0	0	$\frac{1}{3}$	0	0	1	0	0	0	0	0	0	0	0
S:	0	0	$\frac{1}{3}$	0	0	0	0	0	0	$\frac{2}{3}$	0	0	0	$\frac{1}{3}$	$\frac{2}{3}$	0	0
T:	0	0	0	0	0	0	$\frac{2}{3}$	0	0	0	0	0	0	0	$\frac{1}{3}$	0	$\frac{2}{3}$
V:	0	$\frac{1}{3}$	0	0	0	0	0	0	0	0	$\frac{1}{3}$	0	0	0	0	0	0
–:	0	0	$\frac{1}{3}$	0	$\frac{1}{3}$	$\frac{2}{3}$	0	0	0	$\frac{1}{3}$	0	0	0	$\frac{2}{3}$	0	0	$\frac{1}{3}$

（4）通过使用该一级结构谱进行的序列同源性识别：

一级结构谱：

氨基酸残基：			–								V						
残基相同的得分：			$+\frac{1}{3}$								$+\frac{1}{3}$						
氨基酸残基：		V	S		–	–	V			–	A			–	T		–
残基相同的得分：		$+\frac{1}{3}$	$+\frac{1}{3}$		$+\frac{1}{3}$	$+\frac{2}{3}$	$+\frac{1}{3}$			$+\frac{1}{3}$	$+\frac{1}{3}$			$+\frac{2}{3}$	$+\frac{1}{3}$		$+\frac{1}{3}$
氨基酸残基：	P	D	A	C	A	P	T	E	P	S	D	F	V	S	S	D	T
残基相同的得分：	+1	$+\frac{2}{3}$	$+\frac{1}{3}$	+1	$+\frac{2}{3}$	$+\frac{1}{3}$	$+\frac{2}{3}$	+1	+1	$+\frac{2}{3}$	$+\frac{1}{3}$	+1	+1	$+\frac{1}{3}$	$+\frac{2}{3}$	+1	$+\frac{2}{3}$

序列1：F V S F A V A A A S V – F S S A T

得分：$0+\frac{1}{3}+\frac{1}{3}$ $0+\frac{2}{3}$ 0 0 0 $0+\frac{2}{3}+\frac{1}{3}-1$ $0+\frac{1}{3}+\frac{2}{3}$ $0+\frac{2}{3}$ $=+3$

不同源

一级结构谱：

氨基酸残基：			–									V						
残基相同的得分：			$+\frac{1}{3}$									$+\frac{1}{3}$						
氨基酸残基：		V	S		–	–	V				–	A			–	T		–
残基相同的得分：		$+\frac{1}{3}$	$+\frac{1}{3}$		$+\frac{1}{3}$	$+\frac{2}{3}$	$+\frac{1}{3}$				$+\frac{1}{3}$	$+\frac{1}{3}$			$+\frac{2}{3}$	$+\frac{1}{3}$		$+\frac{1}{3}$
氨基酸残基：	P	D	A	C	A	P	T	E	P	–	S	D	F	V	S	S	D	T
残基相同的得分：	+1	$+\frac{2}{3}$	$+\frac{1}{3}$	+1	$+\frac{2}{3}$	$+\frac{1}{3}$	$+\frac{2}{3}$	+1	+1		$+\frac{2}{3}$	$+\frac{1}{3}$	+1+1	$+\frac{1}{3}$	$+\frac{2}{3}$	$+\frac{2}{3}$	+1	$+\frac{2}{3}$

序列2：P F F C E – E E P T T T F V – A D A

得分：+1 0 0+1 $0-\frac{2}{3}$ 0+1+1−1 0 0+1+1$-\frac{2}{3}$ 0+1 0 $=4\frac{2}{3}$

同源

对“得分”中的小数的解释：−1 为序列中的空位（“–”）加上（2/3）（+1/2），其中+1/2 为在序列中的残基相同“–”和在多序列比对中的残基相同“–”，而 2/3 为“–”在多序列比对的给定位置的权重。

习题 22.3

请使用图 22-6 中的示意图，估算如下序列中的哪一些包含 α 螺旋，哪些包含不规则环线，哪些包含 β 串，哪些包含 β 转折。

（1）Gly-Thr-Ser-Val-Thr-Val-Ser-Phe-Thr-Trp-Gly-Gly

（2）Ser-Gly-Gly-Pro-Gln-Glu-Ser-Asn-Gly-Tyr-Arg-Pro

（3）Val-Ala-Gly-Pro-Gly-Ser-Ser-Gly-Pro-Pro-Gly-Val

（4）Gly-Pro-Ala-Glu-Gln-Leu-Leu-Gln-Lys-Met-Arg-Gly

解答:

(1) β 串;(2) 不规则环线;(3) β 转折;(4) α 螺旋

第 23 讲

习题 23.1

假设,在蛋白质结构预测工作中,测定结构的能量时的典型误差为$\pm\sigma$,而且值为 ΔE 或绝对值超过 ΔE 的误差的概率服从高斯定律:$P(\Delta E)\sim\exp\left(-\frac{\Delta E^2}{2\sigma^2}\right)$。

这样的计算能很好地区分能量之差为 10σ 的两个结构:它可以几乎肯定地说,其中哪一个比另一个更稳定。(例如,两个能量之差以标准误差 $\pm\sqrt{2}\sigma$ 进行估算,则在宽为 10σ 的能隙中含有 ≈ 7 个标准误差,即在测定能隙的存在时的错误概率,根据误差概率表,$\sim 10^{-11}\sim e^{-25}$。这个估计也可以这样获得:低能量结构的能量升高 5σ 的误差为 $\sim e^{-25/2}$;高能量结构的能量降低 5σ 的概率也为 $\sim e^{-25/2}$;意味着,这两个事件同时发生的概率为 $\sim e^{-25/2}\cdot e^{-25/2}=e^{-25}$)。

现在假设,我们需要比较被预测得的 M 个($M\gg 1$)结构的能量,而且只有其中一个是稳定的,即具有低能量,而所有其他结构的真实能量要高 10σ。

M 为什么样的数目时,以上述误差进行计算的结果不再可靠,即其中一个高能结构的计算能量等于实际上比低能结构高 10σ 的某个结构的计算能量?

解答:

设低能结构的真实能量等于 0,而高能结构的真实能量等于$+10\sigma$。则在 $M-1\approx M$ 个高能结构中的任一个结构的计算得的能量小于 0 的数学期望值为 $(M-1)\cdot 0.5\cdot\exp[-(10\sigma)^2/(2\sigma^2)]\sim M\exp(-50)$,如果 $M\exp(-50)\leqslant 0$,则某个或某些高能结构的计算能量与真实的低能结构的计算能量发生重叠。

结果:在 $M\leqslant e^{50}\approx 10^{22}\approx 2^{70}$时能可靠地找到低能结构;而由于每个氨基酸有多于两个的可能构象,则这种计算的结果对超过 70 个残基的链会不再可靠。

第 24 讲

习题 24.1

在室温下,免疫球蛋白与某种分子 X 相结合,结合常数为 $K_{ass}=\frac{[Ig-X]_{eq}}{[Ig]_{eq}[X]_{eq}}=10^8$L/mol。假设这种分子有球状外形(即没有内部的自由度),且它可在免疫球蛋白内振动,振幅为 $a=0.5$Å。

请估算该分子与免疫球蛋白的相互作用自由能和该分子与水的相互作用自由能之差 ΔG_0。

解答:

$\Delta G_0\approx -15.5$kcal/mol。

解释：

粒子 X 在溶液中的浓度$[X]_0 = 1/K_{ass}$时，$[Ig-X]/[Ig] = 1$，即免疫球蛋白处于 Ig-X 形态和 Ig 形态的概率相同。这意味着，在这种情形下这两种形态的自由能相同。免疫球蛋白从溶液中抽取粒子时，它做功用于减小粒子可及体积：在溶液中其所占据体积的比率为：$V_0 = 1/[X]_0 = K_{ass}$，而在免疫球蛋白中只有其振动体积：$V_1 = (2a)^3$。相应的功等于 $kT\ln(V_0/V_1)$。对免疫球蛋白的这个功的估算要加上与粒子的附加相互作用(确实，这替换了其结合口袋和粒子结合表面原先与水的相互作用)。粒子-免疫球蛋白的相互作用自由能与它们所失去的与水的相互作用自由能之差 ΔG_0，应该能补偿在减小粒子可及体积上所花费的功：因为 Ig-X 形态与 Ig 形态的自由能是相同的。这意味着，$\Delta G_0 + kT\ln(V_0/V_1) = 0$，最后得到：$\Delta G_0 = -kT\ln[K_{ass}/(2a)^3]$。

数值计算：$K_{ass}/(2a)^3 = (10^8\text{L/mol})/1\text{Å}^3 = (10^8\times10^3\text{cm}^3/\text{mol})/(10^{-8}\text{cm})^3 = [10^{11}/(0.6\times10^{24})]/10^{-24} = 1.7\times10^{11}$。在 $kT = 0.6$kcal/mol 时，有 $\Delta G_0 \approx -15.5$ kcal/mol.

习题 24.2

最简单的酶促反应动力学模型(Michaelis-Menten 模型)把这种反应解释为由自由底物 S 和自由酶 E 可逆地生成中间体 ES，伴随着处于 ES 组分中的底物不可逆地转变为产物 P。而且酶-底物复合体 ES 由自由底物和自由酶生成时的速率为$[E]\cdot[S]\cdot k_1$，分解为自由底物和自由酶的速率为$[ES]\cdot k_2$，而分解为自由产物和自由酶的速率为$[ES]\cdot k_3$(如图所示)。



$$E+S \underset{k_2}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_3} E+P$$

请计算在溶液中产物 P 的生成速率与全部底物($[S]_T\equiv[S]+[ES]$)和全部酶($[E]_T\equiv[E]+[ES]$)之间的依赖关系。(假设，溶液中的底物远多于酶，则该计算可以在定态近似下进行。根据该近似，中间体浓度的变化速率 $d[ES]/dt$ 远比 $d[S]/dt$ (或 $d[P]/dt$)要小，中间体的浓度[ES]可以认为是不变的。)

解答：

在定态近似下，$\frac{d[ES]}{dt} = [E]\cdot[S]\cdot k_1 - [ES]\cdot k_2 - [ES]\cdot k_3$，于是，

$$[ES] = \frac{[E]\cdot[S]}{K_M}, \tag{24.2.1}$$

其中，

$$K_M = (k_2+k_3)/k_1 \tag{24.2.2}$$

称为 Michaelis 常数。

由于溶液中的底物很多，而酶很少，则自由底物的浓度[S]接近于底物的全浓度$[S]_T$(而$[ES]\ll[S]_T$)；而自由酶的浓度会比其全浓度$[E]_T=[E]+[ES]$低得

多。把$[E]=[E]_T-[ES]$代入式(24.2.1),得

$$K_M[ES]=([E]_T-[ES])\cdot[S]_T$$

并进一步得到：

$$[ES]=\frac{[E]_T\cdot[S]_T}{[S]_T+K_M}$$

请注意,在生成酶-底物复合体时,当$[S]<K_M$时只有小半酶参与,当$[S]>K_M$时有多半酶参与。

最后,产物的生成速率为

$$\frac{d[P]}{dt}=[ES]\cdot k_3=k_3\cdot\frac{[E]_T\cdot[S]_T}{[S]_T+K_M} \qquad (24.2.3)$$

习题 24.3

在求解上一题中得到的 Michaelis 常数$K_M=(k_2+k_3)/k_1$与复合体 ES 分解成 E 和 S 的解离常数K_{diss}之间的关系如何？

解答：

复合体 ES 分解成 E 和 S 的解离常数$K_{diss}=[E]\cdot[S/ES]_{eq}$,其中$[ES]_{eq}$为复合体 ES 在既不减少也不增加、不存在底物到产物的转变(即$k_3=0$)时处于与给定的底物和酶的浓度($[E]$和$[S]$)相平衡时的浓度。这种平衡条件可以表示为 $d[ES]/dt=[E]\cdot[S]\cdot k_1-[ES]_{eq}\cdot k_2=0$的形式。于是有

$$k_2/k_1=[E]\cdot[S]/[ES]_{eq}\equiv K_{diss}. \qquad (24.3.1)$$

而 Michaelis 常数可以表示为如下形式：

$$K_M=(k_2+k_3)/k_1=(k_2/k_1)(1+k_3/k_2)=K_{diss}\cdot(1+k_3/k_2), \qquad (24.3.2)$$

式中,k_3/k_2项表示酶-底物复合体分解释放出产物和释放出底物的比例。在$k_3\ll k_2$时,有$K_M\approx K_{diss}$。

习题 24.4

假设底物转变为产物的自发(非催化的)反应以$S\longleftrightarrow S^*\longrightarrow P$(其中$S^*$为底物的过渡态)的机制发生,其活化自由能为$\Delta F_0{}^{\#}\equiv F(S^*)-F(S)$。而在酶的作用下,底物转变为产物的反应以$ES\longleftrightarrow ES^*\longrightarrow EP$的机制发生,但活化自由能较低(由于底物过渡态与酶的相互作用):ΔF_{cat}^{J}HJ$\equiv F(ES^*)-F(ES)=[F(S^*)-F(S)]-\delta$。

(1) 催化速率常数k_3(见习题 24.2)与底物到产物的自发转变速率k_0之比如何？

(2) 酶催化速率(见习题 24.2 的解)与底物到产物的自发反应速率之比如何？

解答：

(1) 由于克服活化能垒的速率可以表示为$k=k_{0,el}\cdot\exp(-\Delta F^{\#}/k_BT)$,其中$k_{0,el}=k_BT/h$为在温度$T$下的热振动频率,而$\Delta F^{\#}$为活化自由能,则$\Delta F_{cat}^{J}$相对于$\Delta F_0{}^{\#}$降低$\delta$,能相对于$k_0$增大$\exp[-(-\delta)/k_BT]$倍,变为$k_3$,即

$$k_3=k_0\cdot\exp(\delta/k_BT)$$

（2）由于从底物到产物的酶催化转变速率为 $k_3 \cdot \frac{[E]_T \cdot [S]_T}{[S]_T + K_M}$，而底物自发转变为产物的速率为 $k_0 \cdot [S]_T$，则相对于 k_0 增大 $\exp[-(-\delta)/k_B T]$ 倍变到 k_3，能把使酶催化过程的速率比自发转变速率增大 $\exp\left(\frac{\delta}{k_B T}\right) \cdot \frac{[E]_T}{[S]_T + K_M}$ 倍。

第 25 讲

习题 25.1

在天然蛋白质球体中，某些粒子处于由其与远程链区段的相互作用而产生的一维势阱内。在粒子偏离阱中心 x 时，粒子的能量 $E=\gamma x^2/2$，且粒子在阱中振动，平均振幅为 $\sqrt{x^2}=0.25Å$，而且粒子横穿过阱的平均时间约为 10^{-9}s。

需要等待在数量级上多长的时间，粒子会偏离阱中心（1）1.0Å；（2）2.0Å；（3）3.0Å？考虑到天然蛋白质在每昼夜内临时完全解折叠一次（见习题 20.3）。

解答：

（1）大约 10^{-6}s；（2）大约半天；（3）大约一昼夜。

解释：

粒子在天然蛋白质中以势能 $E=\gamma x^2/2$ 偏离阱中心 $\sqrt{x^2}$，对应于等于 $kT/2$ 的势能（参见习题 8.5）。这意味着，偏离该阱中心 0.25Å 对应于势能 $kT/2$，偏离 1.0Å 对应于势能 $8kT$，偏离 2.0Å 对应于势能 $32kT$，偏离该阱中心 3.0Å 对应于势能 $72kT$，具有能量 E 的偏离的等待时间为 $\sim\tau\times\exp(-E/kT)$。在 $E=kT/2$ 时的等待时间为 $\sim\tau\times\exp(-1/2)\sim10^{-9}$s，而对任意 E 的等待时间为 $\sim\tau\times\exp(-E/kT)\sim(10^{-9}\text{s})\times\exp(-1/2E/kT)$。于是，在偏离 1.0Å 时为 $\sim2\times10^{-6}$s，在偏离 2.0Å 时为 $\sim5\times10^4\text{s}\approx13\text{h}$，在偏离 3.0Å 时为（按公式计算）$\sim10^{22}\text{s}\sim10^{14}$年。不过，在实际上，偏离 3.0Å 不需要等待 10^{14}年，而只是一昼夜，因为天然蛋白质大约每昼夜完全解折叠一次，而在解折叠的情况下，由与远程链片段相互作用产生的势阱会消失。

习题 25.2

ATP 在水中的解离常数为 $K_{diss}=\frac{[ADP][P]}{[ADP]_{eq}}=5\times10^5$mol/L，其中下标 eq 表示 ATP 在给定的 ADP 和磷酸根 P 浓度下与它们达到平衡时的浓度。

（1）在生理浓度 $[ATP]_{生理}\approx10^{-3}$mol/L，$[ADP]_{生理}\approx10^{-5}$mol/L，$[P]_{生理}\approx10^{-3}$mol/L下，一个 ATP 分子的水解自由能为多少？

（2）在什么样的 ATP 浓度下（在同样的 $[ADP]_{生理}$ 和 $[P]_{生理}$ 下），ATP 的分解会超过其合成？

解答：

(1) 在生理浓度下，$\frac{[ADP]_{生理}[P]_{生理}}{[ATP]_{生理}}=\frac{[10^{-5}mol/L]\cdot[10^{-3}mol/L]}{[10^{-3}mol/L]}\approx 10^{-5}mol/L$。

这意味着在生理条件下 $[ATP]_{生理}=\frac{[ADP]_{生理}[P]_{生理}}{10^{-5}mol/L}$，而在平衡时有 $[ATP]_{eq}=\frac{[ADP]_{生理}[P]_{生理}}{5\times 10^{5}mol/L}$，即 ATP 分子生理条件下的浓度比在 ATP 与 ADP 和 P 的生理浓度达成平衡时的浓度要多 $5\cdot 10^{5}/10^{-5}=5\cdot 10^{10}$ 倍。这种超过量对应着多余的自由能，水解自由能等于 $kT\cdot\ln(5\cdot 10^{10})\approx 25\ kT\approx 15kcal/mol$。

(2) ATP 的分解超过其合成，则（在同样的 $[ADP]_{生理}$ 和 $[P]_{生理}$ 下）ATP 的浓度只有 $[ATP]_{生理}$ 的 $5\cdot 10^{10}$（见上文）之一，即 $\approx(10^{-3}mol/L)/5\cdot 10^{10}\approx 0.2\cdot 10^{-13}mol/L$。

习题 25.3

肌球蛋白头部分解 ATP，所获得的自由能的 50%（15kcal/mol 的 50%）转换为可输出的肌肉功。每次分解导致由肌球蛋白头部相对于肌动蛋白的 5nm 位移引起的一次力量冲击。

(1) 请估算肌球蛋白头部产生多大的力。

(2) 在肌肉收缩时，肌肉末端以多大的速度相互靠近？设肌肉在纵向上有 30 000 个肌节，同时工作的肌球蛋白头部占总数的 1%，而每个头部每秒钟分解 20 个 ATP 分子。

(3) 如果在肌肉中有 10^{-6} 摩尔的肌球蛋白头部，则该肌肉中能产生多大的力？

(4) 该肌肉能产生多大的功率？

(5) 若其他参数（尤其是肌球蛋白头部的数量）不变，如果肌肉在纵向上不是有 30 000 个肌节，而是 10 000 个肌节，那么肌肉的收缩速度、力量和功率如何变化？

解答：

(1) 每一个肌球蛋白头部产生的力等于：

$$\frac{15kcal/mol\times 0.50}{5nm}=\frac{4.2\times 15\times 10^{3}J/mol\times 0.5}{5\times 10^{-9}m}=6\times 10^{12}N/mol=10^{-11}N$$

(2) 由于在每一时刻在 100 个头部中只有一个工作，则在 1/20s 内，即在每个肌球蛋白分解一个 ATP 分子的时间内，在同一个肌动蛋白和肌球蛋白丝的连接中，能够工作的不是一个，而是 100 个头部；这导致这些肌丝的相对位移为 5nm×100=500nm。由于沿着肌节有两个肌动蛋白和肌球蛋白丝的连接[在其右边和左边，图 25-8(a)]，则肌节在 1/20s 内收缩 2 倍的大长度，即 1000nm，而 30 000 个相对行走的肌节收缩的长度为 30 000 000nm=0.03m（也在 1/20s 内）。这意味着，这种肌肉的收缩速度为 0.6m/s。

(3)在 1/20s 内,10^{-6}mol 的肌球蛋白头部分解 10^{-6}mol 的 ATP(其这 50% 的分解自由能转换为功),而且肌肉收缩 0. 03m,意味着,

$$力=\frac{15kcal/mol\times0.50\times10^{-6}mol}{0.03m}=0.25cal/m\approx1J/m=1N$$

(4) 功率=力×速度,即 1N×0. 6m/s=0. 6W

(5) 功率(=能量/时间)无变化,收缩速度降低到 1/3,力变大到 3 倍。

索 引